Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz

Original article

AN EXPLAINABLE DEEP LEARNING FRAMEWORK FOR BREAST CANCER CLASSIFICATION USING EFFICIENTNETV2B0 AND GRAD-CAM.

Jane M. Haj Ali[1,*], Diman Hassan[1]

Department of Computer Science, College of Science, University of Zakho, Zakho, Kurdistan Region, Iraq.

*Corresponding Author Email: jane.136001351@stud.uoz.edu.krd

Received: 30, Jun, 2025

Accepted: 29, Jul, 2025

Published: 13, Jan,2026

ABSTRACT

Breast cancer remains one of the most serious health challenges worldwide, where early and accurate diagnosis can significantly improve patient outcomes. Traditional diagnostic methods often rely heavily on expert interpretation, which may lead to inconsistencies or delays in decision-making. To address this issue, this research provides a deep-learning framework that uses the EfficientNetV2B0 model in combination with Grad-CAM (Gradient-weighted Class Activation Mapping) to provide illustrated explanations to detect breast cancer using ultrasound and MRI datasets. Our method addresses serious challenges such as class imbalance and irrelevant image characteristics by employing SMOTE (Synthetic Minority Over-sampling Technique) oversampling and Region of Interest (ROI) extraction for BUSI (Breast Ultrasound Images) datasets. The Grad-CAM approach improves reliability and transparency by providing visual proof that supports each decision, enabling healthcare professionals to better understand the AI's predictions. Trained and assessed on two different medical imaging datasets, the framework obtained extraordinarily high accuracy (98.97% on BUSI and 99.55% on MRI), along with low prediction error and high reliability. The model is both accurate and understandable, making it ideal for clinical usage. It is also faster and more dependable than current approaches, making it highly beneficial.

KEYWORDS: Deep Learning, Breast Cancer, Explainable Artificial Intelligence (XAI), Region of Interest (ROI), Grad-CAM heatmap.

1. INTRODUCTION

Breast cancer is a primary cause of death among women worldwide, accounting for a significant number of cancer-related fatalities despite remarkable advances in healthcare and screening technologies (Waks & Winer, 2019). Since 2020, the global burden of breast cancer has increased, along with considerable geographic disparities. In 2022, there were an anticipated 2.3 million new cases and 670,000 deaths worldwide. If current patterns continue, yearly consistence is predicted to reach 3.2 million, with a mortality rate of 1.1 million by 2050 (Arnold *et al*., 2022). This incidence is expected to have a disproportionate effect on countries with low Human Development Indexes (HDI), as access to early identification and treatment remains limited. While some high-HDI nations have reduced death rates, many low-HDI nations are witnessing increases, highlighting persistent global disparities in breast cancer outcomes. Among the diversity of imaging modalities available, the

Access this article online

https://doi.org/10.25271/sjuoz.2026.14.1.1659

Printed ISSN 2663-628X;
Electronic ISSN 2663-6298

Science Journal of University of Zakho
Vol. 14, No. 01, pp. 83 –96 January-2026

breast ultrasound is recognized as a cornerstone in breast cancer detection due to its non-invasive nature, real-time imaging abilities, and absence of ionizing radiation, making it a preferred option for repeated examinations (Bhushan *et al*., 2021). Ultrasound shines at distinguishing between solid and cystic tumors, and it's particularly helpful for women with dense breast tissue or when mammograms leave questions unanswered (Burkett & Hanemann, 2016). Yet, ultrasound is just one piece of the puzzle. On the other hand, Magnetic Resonance Imaging (MRI) has become an equally vital tool in the fight against breast cancer. MRI offers incredibly detailed, cross-sectional images of breast tissue, revealing subtle differences that might slip past ultrasound or mammography. Its high sensitivity makes it especially valuable for detecting both primary tumors and additional lesions, which is crucial for patients with dense breasts or those at higher risk (Ma *et al*., 2025). In this way, MRI does not replace ultrasound, but complements it, providing a broader, more nuanced view and empowering clinicians to make treatment decisions with greater confidence. These image techniques work well together; each contributes to the others' strengths, ensuring that patients receive the best possible, comprehensive therapy. On top of that, the two forms of images have distinct and complementary roles in assessing breast cancers. Breast tumors are classified into two types in medical terms: benign and malignant. Benign tumors are often harmless, rarely spread, and can be removed surgically with a low risk of regrowth. Tumors that are malignant, on the other hand, attack nearby body tissues, can spread to other organs, and present an even greater chance of death (Akram *et al*., 2017). Precise tumor classification is critical, as it directly affects treatment methods. For benign cases, this involves careful observation; for malignant situations, it requires aggressive actions such as immunotherapy, radiation therapy, chemotherapy, and surgical removal (Waks & Winer, 2019). The issues with traditional ultrasound image interpretation, including time-consuming analysis, differences among radiologists, and the risk of incorrect diagnosis due to hidden or unclear imaging features, can be resolved through the revolutionary adoption of Artificial Intelligence (AI) in medical imaging. Deep learning, a subset of AI, uses large datasets and powerful neural networks to discover complex patterns in images, often surpassing human experts. Even though it's interesting, the "black box" characteristics of many models have limited the adoption of deep learning in healthcare environments (Rasheed *et al*., 2022). Even while these models work effectively, their absence of openness makes it hard to fully understand the reasoning behind their predictions. Specialists may become suspect as a result of this, because even slight input changes could lead to serious mistakes, weakening trust and responsibility (Hassija *et al*., 2024). These concerns have helped in the creation of Explainable Artificial Intelligence (XAI), a multifaceted area focused on improving the understanding of AI systems by explaining their method of decision-making (Adadi & Berrada, 2018). XAI is especially crucial in healthcare as transparency ensures that AI-powered findings align with clinical expertise and diagnostic standards (Albahri *et al*., 2023). The most prevalent XAI methods are gradient-based, which perform well with complicated nonlinear models (Mersha *et al.*, 2024). Using the EfficientNetV2B0 model (Tan & Le, 2021), a state-of-the-art Convolutional Neural Network (CNN) known for its ability to balance computational efficiency and classification prowess, this study presents an advanced AI framework for classifying breast ultrasound images as benign or malignant. EfficientNetV2B0 optimizes performance through a compound scaling approach and Fused-MBConv layers, making it ideal for resource-constrained contexts like clinical settings. Meanwhile, the Synthetic Minority Over-sampling Technique (SMOTE) addresses the problem of class imbalance in medical image datasets, where benign cases are often more numerous than malignant cases (Chawla *et al.*, 2002). This approach generates artificial samples from the minority (malignant) subset, ensuring balanced training while reducing bias towards the majority class. Besides, by highlighting features such as irregular boundaries or echogenic textures, the Region of Interest (ROI) for BUSI dataset extraction reduces background noise and improves diagnosis by focusing the model on tumor-specific areas. Our approach relies on Gradient-weighted Class Activation Mapping (Grad-CAM), an XAI technique that produces visual heatmaps showing the image regions most valuable to the model's predictions.(Selvaraju *et al.*, 2020). By bridging the gap between clinical understanding and computational outputs, those heatmaps enable specialists to verify that the model's highlights correspond to established diagnoses, thereby fostering trust and acceptance of clinical methods. To test our model, the Breast Ultrasound Images BUSI dataset and the Breast Cancer Patients MRI dataset were employed. The primary findings of this research are as follows:

1. The proposed method uses EfficientNetV2B0 to achieve a remarkable result 99.55% for test accuracy on the MRI dataset and 98.97% on the BUSI dataset, going over many state-of-the-art techniques.

2. By integrating SMOTE and ROI extraction, the model's sensitivity and specificity is improved, effectively addressing class imbalance while sharpening its focus on key diagnostic features.

3. The use of the XAI technique Grad-CAM in this work offers clear visual insights into the model's decision-making, bridging AI predictions with radiological expertise to boost its practical value in clinical settings.

The structure of the paper is structured as follows: Section 2 presents related work. Materials and methods,

which outline the dataset, preprocessing steps, model architecture, method, and evaluation metrics, are presented in Section **3**. Furthermore, the results of this study and their discussion are reported in Section 4, and finally, the conclusion, limitation, and future work are presented in Section **5**.

RELATED WORKS

Over the past few years, researchers have been exploring novel ways to use artificial intelligence to enhance breast cancer screening using different types of images. These investigations aim to develop powerful computer models capable of analyzing images, identifying tumors, and determining whether tumors are cancerous. These efforts, which combine cutting-edge technology with medical imaging, aim to make diagnoses faster, more accurate, and easier for doctors to utilize. While each strategy has advantages and disadvantages, they all demonstrate how AI is changing the way it identifies and comprehends breast cancer, offering the promise of improved outcomes in the future.

In this context, (Vigil *et al.*, 2022) developed a deep learning model to segment breast lesions and extract radiomic characteristics from ultrasound pictures. The model uses a convolutional autoencoder and a contracting-expanding architecture to reduce high-dimensional radiomic data. The model's dual functionality reduced the need for separate pipelines, thereby increasing efficiency. A random forest classifier attained a cross-validated accuracy of 78.5% in differentiating between malignant and benign cases.

Building on the theme of integrated workflows, (Podda *et al.*, 2022) introduced an automated deep learning system that used a mix of CNN models to classify and segment breast ultrasound images. By combining models such as ResNet50, InceptionV3, and Xception via soft voting, it achieved 91% classification accuracy. Segmentation was handled using U-Net variants, with a method that refined the masks over time, achieving a 82% Dice score. Overall, it outperformed individual models and existing state-of-the-art methods.

In addition, advancing multi-modal integration, the researcher in (Pathan *et al.*, 2022) utilized a multi-headed convolutional neural network (CNN) to classify breast cancer employing ultrasound images from the BUSI dataset. The program learned raw and masked images separately before merging them to form a lightweight model. The limitations included a small sample size, limited computational resources, and the danger of overfitting. The technique achieved 92.31% accuracy (±2), outperforming single models (78.97% raw, 81.02% masked pictures), and reducing misclassification, especially in malignant cases.

The researchers in (Cruz-Ramos *et al.*, 2023)employed deep learning to create a computer-aided diagnosis system for categorizing benign and malignant breast cancers. The system extracted deep features using DenseNet-201, in addition to bespoke features such as HOG, ULBP, and shape descriptors. Feature selection was done using genetic algorithms and mutual information, and classification was done with XGBoost, AdaBoost, and MLP. The fusion strategy outperformed previous methods, increasing classification accuracy by 97.6%.

The authors in (Zhang *et al.*, 2023) proposed a semantic-aware transformer (SaTransformer) for unified breast cancer classification and segmentation. The method used an encoder-decoder architecture based on the U-Net and a DAM to reduce computational complexity. Among the challenges faced were memory overhead, task interactions, handling imprecise tumor boundaries, and low signal-to-noise ratios. The SaTransformer performed well on the BUSI (97.97% accuracy, DSC: 86.34%) and UDIAT datasets, improved feature representation, and reduced computational costs.

The AEGANB3 method, suggested by (Luong *et al.*, 2024), combines a self-attention mechanism for breast cancer detection with a Deep Convolutional Generative Adversarial Network (DCGAN) to address data shortage issues. DCGAN was used for data augmentation, while EfficientNetB3 was employed for transfer learning and fine-tuning. The main outcome was the self-attention technique's improved feature extraction and 98.01% classification accuracy.

However, according to (Sahu *et al.*, 2024)Strategies from controlled high-performance systems were combined of develop a self-learning breast cancer diagnosis system. The researchers merged three transfer learning models: AlexNet, ResNet, and MobileNetV2. To improve image quality prior to classification, a Gaussian-Laplacian preprocessing was used. The method had a high accuracy of 97.75% for detecting malignant tumors.

Based on the developments in (Jabeen *et al.*, 2024)The authors suggested a deep learning architecture for the classification of breast cancer tumors using ultrasound images. The structure includes a combination of EfficientNet-b0 with a gated recurrent unit (GRU) and modified ResNet-18 with multi-head self-attention. Performing data augmentation, transfer learning, with a novel cuckoo search-based feature selection technique merged with standard error mean computation. Feature fusion used a zero-padding maximum correlation coefficient technique, while Grad-CAM supplied explainability; the framework obtained 98.4% accuracy.

The authors in (Nasir *et al.*, 2022) developed a breast cancer detection model that fine-tunes a pretrained AlexNet neural network using MRI images. To adapt AlexNet for distinguishing between healthy and cancerous breast tissue, they modified the first and last three layers of the network. Since labeled MRI data is often limited, they used transfer learning to overcome this challenge. Their

approach proved highly effective, achieving a test accuracy of 98.1% and a sensitivity of 99%.

Many state-of-the-art methods in prior work do not incorporate explainable AI (XAI) techniques, but only two, Luong *et al.* (2024) and Jabeen *et al.* (2024), explicitly applied them, as highlighted in Tables 3 and 4. This underlines a gap in interpretability across much of the existing literature. In response, our approach leverages Grad-CAM to enhance model transparency.

2. MATERIALS AND METHODS

In this section, an advanced deep learning framework is presented to classify breast cancer images as benign or malignant. The suggested method used the BUSI dataset and MRI dataset to conduct the experimental procedures, as shown in Figure 1. It begins with two different datasets as input images, which are preprocessed by resizing to 299 × 299 pixels. One of the datasets, BUSI, uses ROI to target tumor-specific regions, and the resulting images are then sent to the EfficientNetV2B0 model, which performs classification. We assess the model's predictions using Grad-CAM, which produces a heatmap highlighting the image regions most influential in the model's decision. By examining these regions, clinicians could have a deeper understanding of the model's decision-making procedure.

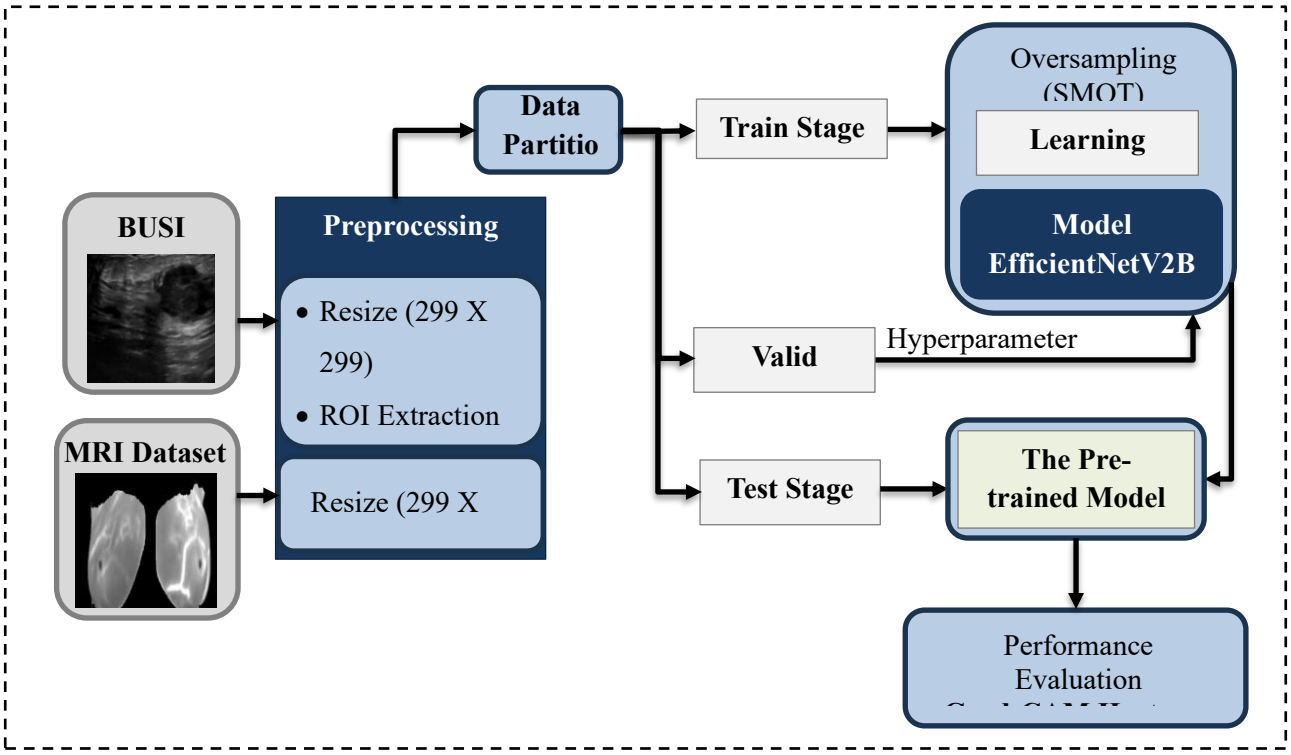

Figure 1: Schematic Diagram of The Proposed Model

Datasets Description:

This work uses two publicly accessible datasets for breast cancer identification: the Breast Ultrasound Images Dataset (BUSI) and the Breast MRI dataset. BUSI (Al-Dhabyani *et al.*, 2020) is a publicly available benchmark dataset designed for breast cancer diagnosis and classification. The data were collected in 2018 at Baheya Hospital in Cairo, Egypt, to support early detection and treatment of breast cancer in women. It included 600 patients varying in age from 25 to 75 years, and the BUSI has 1,578 ultrasound images split into three different subsets: normal, malignant, and benign, each of which has a segmentation mask that defines the borders of the lesions.

The second dataset employed is the publicly available "Breast Cancer Patients MRI's" dataset published on Kaggle (Uzair Khan, 2021), which can be retrieved from https://www.kaggle.com/uzairkhan45/breast-cancer-patients-mris. It is released under a CC0 license, indicating public domain, but this only confirms it is an aggregated set of images from public sources. The dataset contains 1,480 breast MRI images divided into two types: Healthy (Benign) and Sick (Malignant). The dataset is a great resource for improving automated breast cancer identification and diagnosis using MRI images.

Datasets Splitting:

The BUSI dataset and Breast Cancer Patients MRI dataset have both been divided into three distinct subsets: training, validation, and testing. The training set contains 70% of the total data, while the remaining 30% was evenly split between validation and testing subsets, with 15% assigned to validation and 15% to testing. The decision of allocate 70% of the data to training was informed by empirical validation. We experimented with

various train-validation-test split ratios (e.g., 60-40, 80-20), and found that the 70%-30% split consistently yielded the best performance in terms of model generalization and stability on both datasets. This suggests that allocating a larger portion of the data to training helps the model learn more robust feature representations. Given the relatively limited size of both the BUSI and Breast Cancer MRI datasets, using 70% of the data for training is appropriate to maximize the available information for the learning process, while still preserving sufficient data in the validation and test sets for reliable evaluation. Table 1 and 2 shows the distribution of samples for BUSI and MR, respectively.

Table 1: Class-Wise Distribution of Samples Across Training, Validation, And Test Sets for BUSI Dataset

Dataset	Class	Images	Training Set	Validation Set	Test Set
BUSI	Benign	437	305	66	66
	Malignant	210	147	32	31
	Total	**647**	**452**	**98**	**97**

Table 2: Class-Wise Distribution of Samples Across Training, Validation, and Test Sets for The Breast Cancer Patients MRI **Dataset**

Dataset	Class	Images	Training Set	Validation Set	Test Set
MRI	Healthy	740	518	111	111
	Sick	740	518	111	111
	Total	**1480**	**1036**	**222**	**222**

Data Preprocessing:

To facilitate efficient deep learning analysis of breast ultrasound images, the BUSI dataset was first prepared by excluding normal cases, focusing on distinguishing malignant from benign lesions. Then, to highlight tumor locations, the entire dataset was preprocessed by downscaling images to a 299×299-pixel resolution, extracting ROIs using masks, resizing and normalizing the masks, and proportionally scaling them to match the original image size. This same process was applied to the Breast Cancer Patients MRI dataset. While explicit pixel normalization of the raw images was omitted due to the built-in normalization layer in the EfficientNetV2B0 model, the naturally high contrast between tumors and surrounding glandular tissue served as a key visual cue during training. To further enhance model performance, class imbalance was addressed using SMOTE, by flattening each image of size 299×299×3 into a one-dimensional vector (268,203 features) and generating synthetic samples through interpolation in this high-dimensional pixel space. Although this approach does not preserve the spatial structure inherent in image data, it enables the creation of new minority class instances without duplicating existing ones, helping the model learn more effectively and reduce prediction bias, thereby enhancing the diversity of training examples. This non-traditional application of SMOTE was motivated by practical constraints and the need for a reproducible oversampling baseline, as it has been shown to improve model performance in imbalanced settings (Chawla *et al.*, 2002). The use of SMOTE with CNNs and pre-trained models has been supported by several studies. It has been applied effectively to brain MRI (Rajaan *et al.*, 2024) and breast cancer imaging (Joloudari *et al.*, 2023), confirming its adaptability for improving learning from imbalanced medical image data. For the BUSI dataset, SMOTE was applied during training, increasing the number of training images to 610; however, it did not affect the MRI dataset. This entire preprocessing chain aimed to improve input data quality without compromising anatomically meaningful texture patterns, which are crucial for disease classification.

Region of Interest (ROI): ROI, a key idea in image processing, computer vision, and medical imaging, is a specific group of data chosen for a specific study by identifying important areas rich in information, such as tumors in medical imaging (Wang, 2001). In medical imaging, ROI typically highlights abnormal lesions or tumors identified through MRI, mammography, or ultrasound (Nieto-Castanon *et al.*, 2003). Accurate ROI segmentation is crucial for distinguishing benign from malignant tumors, aiding diagnosis and early detection (Krithiga & Geetha, 2021). In this study, the ROI was used to focus the model's attention on tumor areas in breast ultrasound BUSI only. By applying masks to isolate and resize these regions, we preserved important details while minimizing background noise, improving the

model's ability to differentiate between benign and malignant cases. Figure 2 illustrates the ROI extraction process used to guide training for BUSI only. In the MRI dataset, we did not perform ROI extraction because segmentation masks were unavailable.

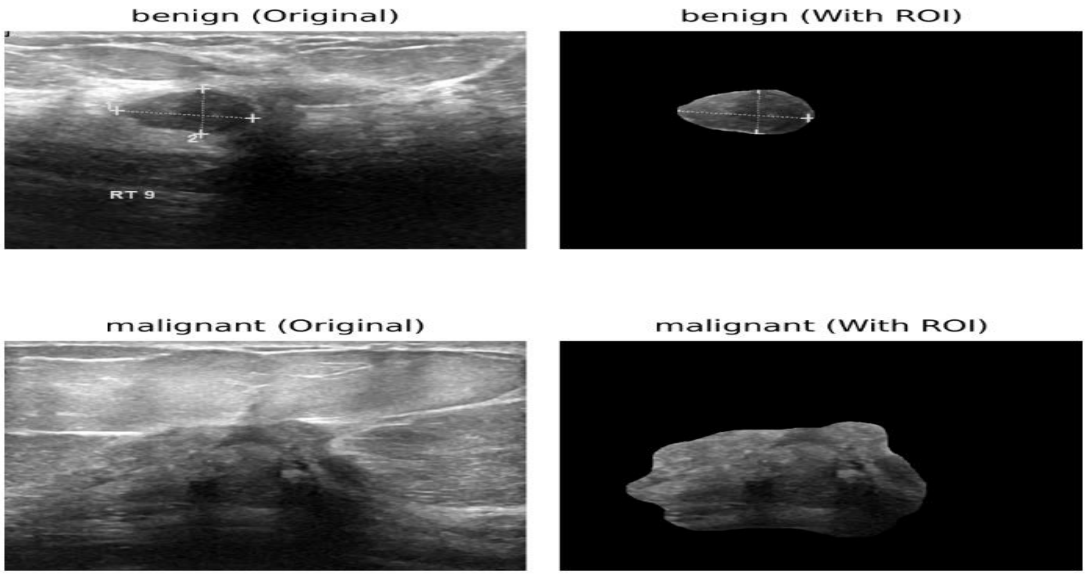

Figure 2: The ROI Extraction Process Used to Guide Training

XAI Grad-CAM Technique: Grad-CAM (Selvaraju *et al.*, 2020) is an XAI approach that enhances understanding of CNNs by providing visual explanations. It is versatile and applicable to a wide range of use applications, such as visual question answering and picture categorization. Grad-CAM can identify model faults, detect bias, and enhance adversarial resilience. Therefore, it improves visualization and produces high-resolution data in medical imaging when used with Guided Backpropagation. Empirical tests have demonstrated its utility for weakly supervised localization and trustworthy assessment, both of which are critical for model validity. Our proposed model's decision-making process was interpreted using Grad-CAM visualizations. Figure 3 displays the final heatmap generated from the model's output, which highlights locations of relevance that impact the model's predictions for the BUSI dataset, whereas Figure 4 presents the highlighted locations of relevance that impact the model's predictions for the Breast Cancer Patients MRI dataset.

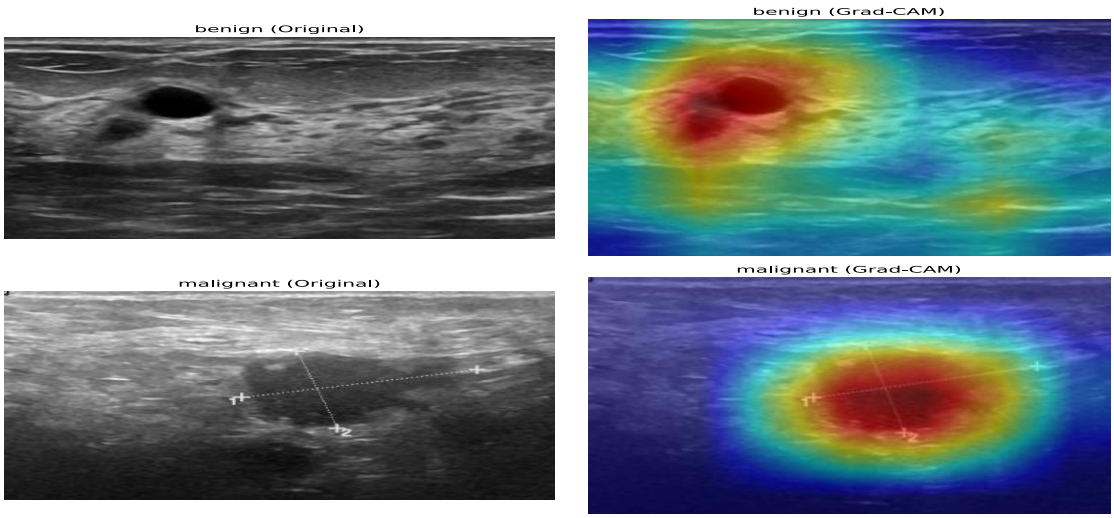

Figure 3: Grad-CAM Visualization for Model Interpretability BUSI Dataset

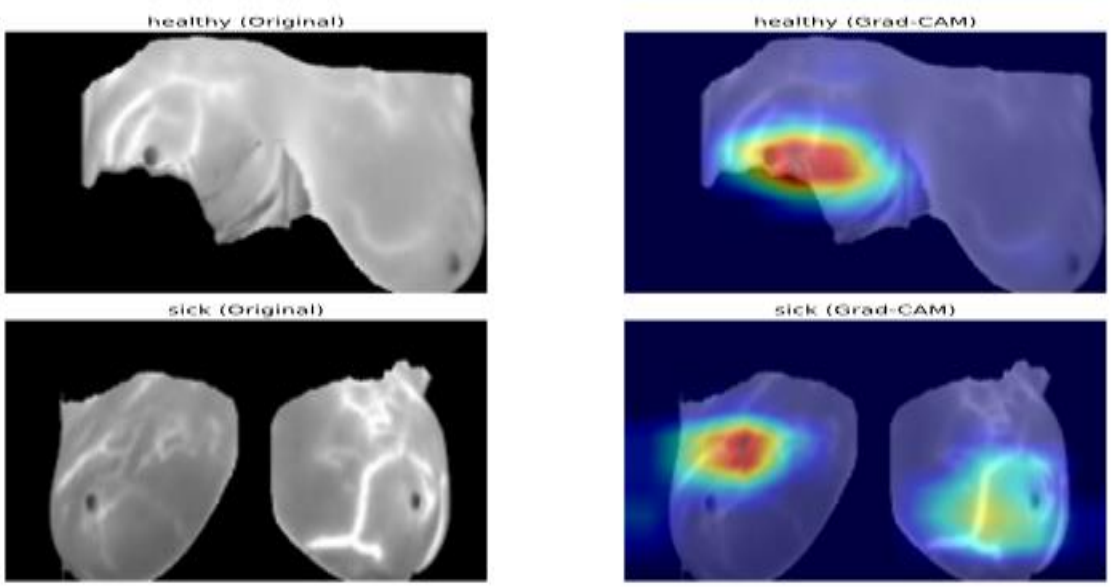

Figure 4: Grad-CAM Visualization for Model Interpretability Breast Cancer Patients MRI Dataset

Model Used in This Study: EfficientNetV2B0:

In this study, the proposed method used EfficientNetV2B0 pre-trained model (Tan & Le, 2021) which is a compact and efficient variant within the EfficientNetV2 family that redefines the balance between model size, training speed, and accuracy. This model is developed by combining a progressive learning methodology with a state-of-the-art training-aware neural architecture search. The adoption of Fused-MBConv operations in the early layers significantly enhances the effectiveness of training by replacing traditional depth wise convolutions, with 7.4 million parameters and 0.7 billion FLOPs, EfficientNetV2B0 defeats bigger models like ImageNet (Tan & Le, 2021). In some situations, it also offers a notable improvement in inference speed, more than two times faster than earlier small-scale models, claiming that the combo of efficiency and power makes it excellent for limited resource usage, such as mobile or edge devices, without losing precision. In addition, the first normalization layer is more than just a preprocessing step; it is a key component of EfficientNetV2B0's efficiency accuracy trade-off, showing the design mindset for execution in limited situations. The model was selected for its strong performance in transfer learning scenarios, particularly where data is limited or imbalanced, as is common in medical imaging. In our implementation, the ImageNet-pretrained EfficientNetV2B0 was fine-tuned for classification tasks involving breast ultrasound and MRI images.

The model was initialized with pretrained ImageNet weights and modified by removing its original classification head (include top=False). Input images of size 299×299×3 were fed into the network and passed through the EfficientNetV2B0 backbone. A Global Average Pooling layer was applied to reduce the spatial dimensions, resulting in a 1280-dimensional feature vector. This vector was then passed to a dense output layer with 2 units and a softmax activation function to enable multi-class probability prediction (i.e., binary classification). We fine-tuned it end-to-end to adapt the pretrained features to the breast cancer imaging task. The model architecture is illustrated in Figure 5, which shows the flow from input to output.

Figure 5: The Block Diagram of The EfficientNetV2B0 Architecture

The choice of EfficientNetV2B0 was motivated by its strong generalization ability in transfer learning tasks, particularly under conditions of limited or imbalanced medical imaging datasets such as breast ultrasound and MRI. The use of global average pooling and a lightweight dense classifier enhances generalization, and the final model supports efficient deployment in clinical and resource-constrained environments.

Evaluation Metrics:

The suggested model's performance in this study for classifying the breast cancer dataset BUSI is evaluated using recall, accuracy, precision, and F1 Score (Jeni *et al.*, 2013) has computed, which are shown as follows:

$$\textbf{Accuracy (ACC)} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

$$\textbf{Recall (Sensitivity)} = \frac{TP}{(TP + FN)} \quad (2)$$

$$\textbf{Specificity} = \frac{TN}{(FP + TN)} \quad (3)$$

$$\textbf{Precision} = \frac{TP}{(TP + FP)} \quad (4)$$

$$\textbf{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (5)$$

The ROC curve shows the false-positive rate (x-axis) and true-positive rate (y-axis)

True positive (TP) samples are those that were correctly predicted as malignant, whereas false positive (FP) samples are those that were forecasted as malignant but were really benign. True negative (TN) denotes correctly predicted benign samples, whereas false negative (FN) denotes predicted benign samples that are really malignant.

EXPERIMENTS AND RESULTS

In this part of the study, the experimental results of applying the EfficientNetV2B0 model to the BUSI dataset and the Breast Cancer Patients MRI dataset, which achieved high performance in breast cancer diagnosis, are described. The accuracy measure, sensitivity rate, precision rate, F1-Score, AUC, and computing time (seconds) are used to calculate each classifier's performance. For the training of the deep models, we divide the dataset into a ratio of 70:15:15. This means that 70% of the images of each class have been utilized for the training of the models, and the remaining 30% used 15% for validation and test 15%. In addition, several hyperparameters have been used to train deep models. The key hyperparameters used in the experiments are summarized in Table 3. The entire experimental process was conducted using Kaggle's free GPU environment (NVIDIA GPU P100), using Python and TensorFlow libraries, and all of these hyperparameters were selected after extensive experimentation. Various combinations were tested, and this configuration consistently delivered the best performance in terms of training stability, convergence, and validation accuracy. While many values could have been used, these settings struck a strong balance between learning efficiency and generalization. By sharing this matrix, we aim to demonstrate the importance of careful tuning and provide transparency for reproducibility.

Table 3: The Key Hyperparameters Used in The Experiments

Hyperparameter	Value
Learning Rate	0.0001
Optimizer	Nadam
Loss Function	Sparse Categorical Cross-entropy
Number of Epochs	22
Mini-batch Size	32
Momentum	(Nadam does not use explicit momentum)
Environment	Kaggle (NVIDIA GPU P100)
Libraries	Python + TensorFlow

3. RESULTS

The proposed framework demonstrated high computational efficiency and strong performance on both the BUSI and MRI datasets. The BUSI dataset was first prepared by excluding normal cases to focus on distinguishing malignant from benign lesions. The total training time was approximately 121.36 seconds, with an average inference time of 0.09 seconds per image for the BUSI dataset. The model achieved test accuracy of 98.97% and an ROC-AUC score of 99.95%, indicating perfect discrimination between benign and malignant breast tumors. When evaluated on the Breast MRI dataset, the framework continued to perform exceptionally well. Training was completed in just 4 minutes and 23 seconds, and the evaluation phase took only 2.26 seconds. The model achieved a test accuracy of 99.55% and an ROC-AUC score of 99.9, confirming its ability to effectively distinguish between healthy and cancerous cases in MRI scans. These results highlight the robustness, speed, and accuracy of the proposed framework in classifying breast cancer using both ultrasound and MRI imaging modalities. Our models Consistency
existing approaches in terms of evaluation metrics, as shown in Tables 4 and 5.

Table 4: Performance Comparison with Existing Models and MRI Dataset.

Reference	Accuracy	F1-Score	Sensitivity	Specificity	XAI Tool	Dataset class
(Nasir *et al.*, 2022)	98.1%	98.1%	99%	97.1%	-	B-M
The proposed method	**99.55%**	**99.54%**	**99.54%**	**99.09%**	**Grad-CAM**	**B-M**

Table 5: Performance Comparison with Existing Models and Busi Dataset

Reference	Accuracy	F1-Score	Sensitivity	Specificity	XAI Tool	Dataset class
(Vigil *et al.*, 2022)	78.5%	-	-	-	-	B-M
(Podda *et al.*, 2022)	91.14%	91.14%	-	-	-	B-M-N
(Pathan *et al.*, 2022)	92.31%	93%	-	-	-	B-M-N
(Cruz-Ramos *et al.*, 2023)	96.10%	96%	96%	96%	-	B-M-N
(Zhang *et al.*, 2023)	97.97%	98.20%	98.31%	93.34%	-	B-M-N
(Luong *et al.*, 2024)	98.01%	98.10%	-	-	Grad-CAM	B-M-N
(Sahu *et al.*, 2024)	96.92%	97.70%	98.08%	94.62%	-	B-M-N
(Jabeen *et al.*, 2024)	98.4%	98.39%	98.43%	-	Grad-CAM	B-M-N
The proposed method	**99.15%**	**99.23%**	**99.49%**	**99.6%**	**Grad-CAM**	**B-M-N**
The proposed method	**98.97%**	**98.82%**	**99.24%**	**98.48%**	**Grad-CAM**	**B-M**

In our experiments, the normal class was not included in the BUSI dataset because its corresponding mask is completely black (i.e., all pixel values are zero). When the ROI extraction method by multiplying the image with its mask is applied, the result is a fully black image with no useful features. As seen in Table 6, the maximum accuracy is obtained when the normal class is included and ROI is applied across all examples; however, this was misleading because the model was not genuinely learning significant patterns, but rather
detecting black images. We removed the normal class to ensure the model learned from informative, valid inputs.

Table 6: BUSI Dataset Accuracy Comparison with and Without (ROI and Normal)

	With ROI	Without ROI
Without Normal	98.97%	86.6%
With Normal	99.15%	93.16%

To address class imbalance in the BUSI ultrasound dataset benign and malignant only, we applied the Synthetic Minority Over-sampling Technique (SMOTE), which resulted in improved classification performance. As shown in Table 7, all key metrics, including accuracy, precision, recall, and F1-score, showed measurable improvement after applying SMOTE. For the MRI dataset, however, SMOTE yielded no significant change, as the class distribution was more balanced and the baseline performance was already saturated.

Table 7: The Effect of SMOTE on BUSI Dataset.

Metric	BUSI Without SMOTE	BUSI With SMOTE
Accuracy	97.94%	98.97%
Precision	97.63%	98.44%
Recall	97.63%	99.24%
F1-score	97.63%	98.82%

As mentioned previously, the MRI dataset had an ROC-AUC of 99.9 for both classes, and the BUSI dataset had an ROC-AUC score of 99.95 for both classes, as shown in Figures 6 and 7, respectively.

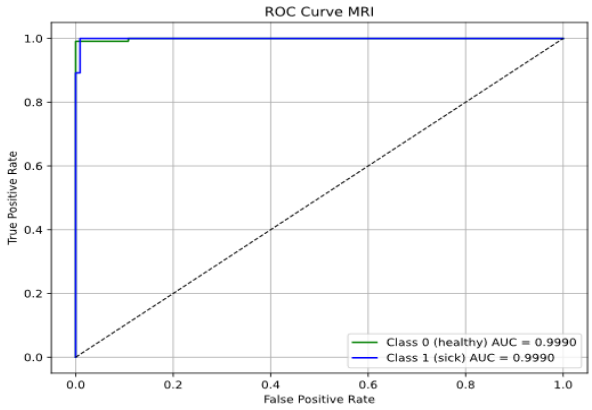

Figure 6: ROC-AUC Using MRI Dataset.

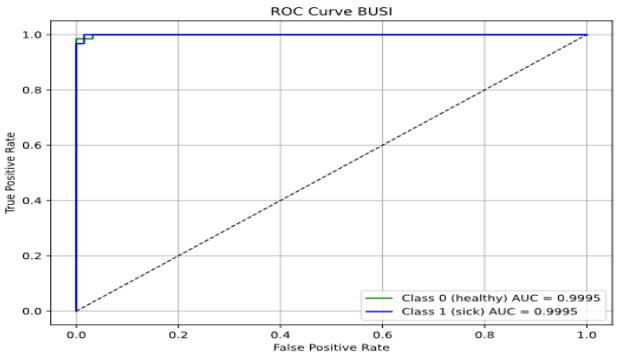

Figure 7: ROC-AUC Using BUSI Dataset

Figure 8 present the confusion matrix of the proposed model using BUSI. The confusion matrix for the BUSI dataset shows that the model performed exceptionally well in distinguishing between benign and malignant cases, with almost no errors. This indicates that the model can confidently detect critical conditions with high precision. The clear separation of predictions suggests strong reliability for medical diagnosis tasks.

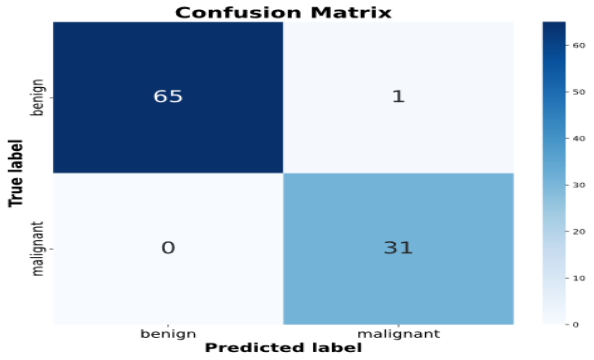

Figure 8: Confusion Matrix of The Classification Model for the BUSI Dataset

Figure 9 presents the confusion matrix of the proposed model using MRI. The MRI dataset results reveal nearly perfect classification, with the model correctly identifying both healthy and sick cases with remarkable consistency. Such performance reflects the model's robustness and its ability to generalize well across different medical imaging data. It reinforces trust in the model's practical deployment for real-world clinical use.

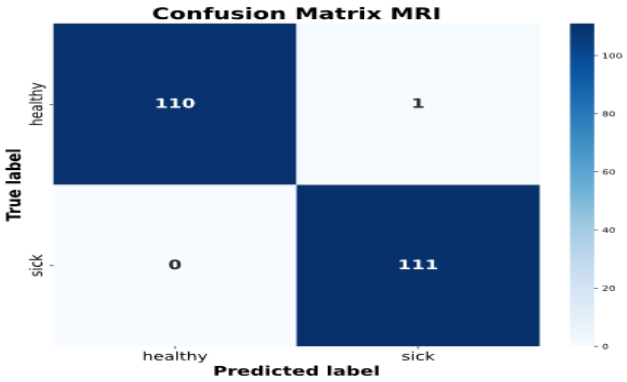

Figure 9: Confusion Matrix of The Classification Model for the MRI Dataset

Discussion and Analysis:

The study introduced a powerful and efficient deep learning framework for breast cancer classification, using both ultrasound and MRI data. It demonstrates that our EfficientNetV2B0-based model, paired with SMOTE, ROI (for BUSI), and Grad-CAM, is an effective tool for diagnosing breast cancer. The high accuracy of 98.97% for BUSI and 99.55% for MRI, along with low error rates and fast processing, demonstrates its reliability and efficiency. For the BUSI dataset, ROI boosted precision by 5.8% (as seen in Figure 10) by focusing on tumor-specific features and cutting out distractions. SMOTE made the model more sensitive to rare cancer cases, while Grad-CAM built trust by showing doctors exactly what the AI was focusing on.

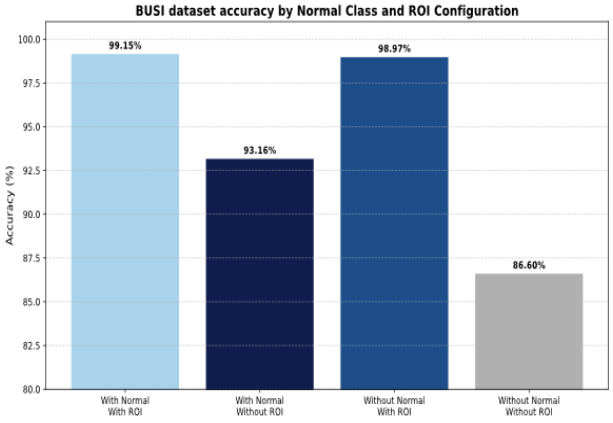

Figure 10: BUSI Dataset Accuracy Comparison Between Models Trained with And Without Region of Interest (ROI) Technique, Under Two Situations with and Without the Normal Class.

The normal class was excluded from the data set because it showed no pathological results and was therefore irrelevant for distinguishing between benign and malignant tumors. Furthermore, in the case of using an ROI, normal images are poorer because their masks are literally empty (zero values). When element-wise multiplication is applied with such a mask, the entire image is suppressed, rendering it unsuitable for training. Moreover, normal images are less effective due to their zero-valued masks, which lead to suppression. This outcome validates the effectiveness of our preprocessing approach, which mitigates the suppression problem caused by zero-valued masks in normal images.

The application of SMOTE was particularly effective for the BUSI ultrasound dataset, where it contributed to measurable gains in classification performance. Improvements in accuracy, precision, recall, and F1-score were observed, confirming that addressing class imbalance through synthetic oversampling improved the model's ability to detect patterns in the minority class. In contrast, the MRI dataset showed negligible improvement with SMOTE, likely due to its more balanced class distribution and the model's ability to learn discriminative features without additional synthetic samples. These results highlight the importance of tailoring preprocessing strategies to the characteristics of each dataset.

However, using the BUSI and MRI datasets helps generalize to different imaging procedures, using ROI for BUSI only, while the MRI dataset did not apply ROI extraction because segmentation masks were not available, making our study suitable for both types of data, with and without segmentation masks. The use of Region of Interest (ROI) segmentation masks was applied to datasets with ground-truth annotations, enabling us to isolate and emphasize diagnostically relevant areas during preprocessing. However, our method was also designed to operate effectively on datasets without segmentation masks by processing the full image. This dual capability demonstrates the flexibility of our approach, allowing it to generalize well across diverse clinical scenarios, both with and without ROI annotations, while maintaining strong classification performance**.** While SMOTE and ROI reduce dataset-specific biases, external validation across many institutional datasets is required to ensure broader application. Furthermore, Grad-CAM improves interpretability; its heatmaps require radiological expertise for accurate clinical translation. The BUSI dataset might not reflect all real-world ultrasound variations (like different patient demographics or imaging setups), thus, more diverse breast cancer data is needed, like an MRI dataset. Although Grad-CAM is helpful for interpreting the heatmaps of the outcomes, it is not considered a replacement for clinicians. The Grad-CAM plays as a supporting technique for them. Figure 11 shows examples from both datasets

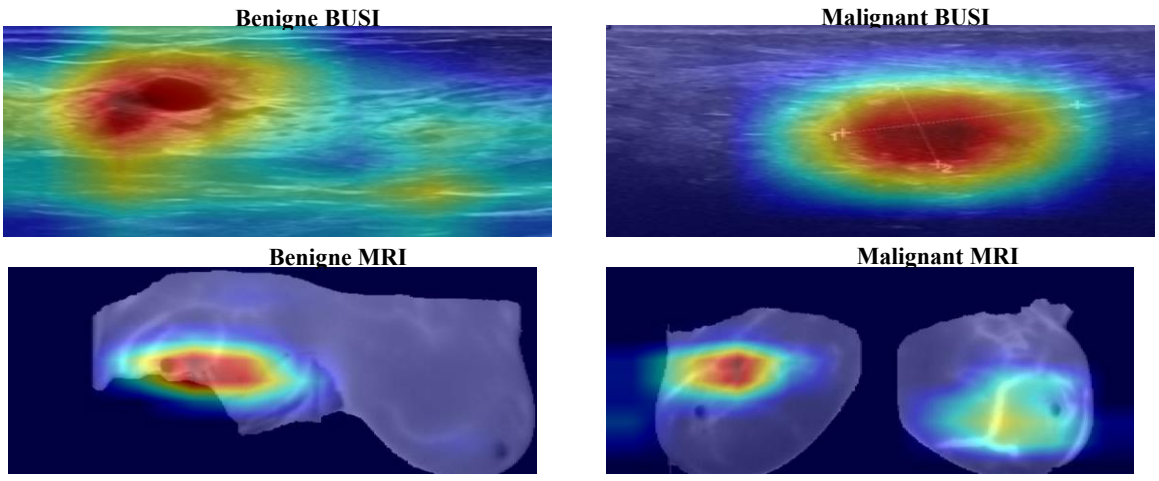

Figure 11: Grad-CAM Heatmap Outcome for Both Datasets.

CONCLUSION

This work demonstrates the effectiveness of an EfficientNetV2B0-based deep learning model in accurately diagnosing breast cancer using both ultrasound and MRI images. The model achieved a test accuracy of 98.97% on the BUSI dataset and 99.55% on the MRI dataset, with ROC-AUC scores of 99.95 and 99.9, respectively. These results were made possible by integrating three core strategies: SMOTE for addressing class imbalance, ROI extraction was used only for the BUSI dataset to improve image focus and clarity, and Grad-CAM for enhancing interpretability by linking predictions to meaningful visual cues.

The model consistently produced minimal prediction errors, indicating robustness and clinical reliability. Using transfer learning and mindful preprocessing, this technique provides a quick, accurate, and explainable detection tool for breast cancer screening that is well-suited for use in real-world medical settings, including resource-constrained settings. This technology provides a highly accurate, interpretable, and strong solution for breast cancer detection, ready to help specialists in real-world settings with precision and transparency.

Limitations and Future Direction:

While existing studies have demonstrated strong performance in breast ultrasound classification, certain limitations remain. Many rely on limited ultrasound datasets that the Model is not tested on Doppler/elastography, test them which could provide greater diagnostic value and improve robustness in real-world clinical environments. Moreover, explainable AI (XAI) techniques are rarely employed only Luong *et al.* (2024) and Jabeen *et al.* (2024) explicitly applied them, as shown in Table 5. In response, our approach integrates Grad-CAM to enhance transparency, with future work exploring additional XAI methods like LIME and SHAP for broader interpretability. Planned directions also include evaluating the system on larger, more diverse datasets, integrating it into real-time ultrasound platforms, and potentially combining it with mammography for a more comprehensive diagnostic pipeline. Clinical validation and the expansion to multi-modal imaging will be essential to ensure the generalizability and practical impact of this work in breast cancer screening.

Acknowledgement:

This publication's work is supported by the University of Zakho, Kurdistan Region, Iraq.

Ethical Statement:

This study has used existing online breast cancer datasets. No experiments have been done to humans.

Author Contributions:

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work J. M. H. A., Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing Original Draft, Writing Review & Editing. D. H., Supervision, Guidance, Writing Review & Editing.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Akram, M., Iqbal, M., Daniyal, M., & Khan, A. U. (2017). Awareness and current knowledge of breast cancer. *Biological Research*, *50*, 1–23. https://doi.org/10.1186/s40659-017-0140-9

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., & Salhi, A. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, *96*, 156–191. https://doi.org/10.1016/j.inffus.2023.03.008

Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, *28*. https://doi.org/10.1016/j.dib.2019.104863

Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast*, *66*, 15–23. https://doi.org/10.1016/j.breast.2022.08.010

Bhushan, A., Gonsalves, A., & Menon, J. U. (2021). Current state of breast cancer diagnosis, treatment, and theranostics. *Pharmaceutics*, *13*(5), 723. https://doi.org/10.3390/pharmaceutics13050723

Burkett, B. J., & Hanemann, C. W. (2016). A review of supplemental screening ultrasound for breast cancer: certain populations of women with dense breast tissue may benefit. *Academic Radiology*, *23*(12), 1604–1609. https://doi.org/10.1016/j.acra.2016.05.017

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cruz-Ramos, C., García-Avila, O., Almaraz-Damian, J. A., Ponomaryov, V., Reyes-Reyes, R., & Sadovnychiy, S. (2023). Benign and Malignant Breast Tumor Classification in Ultrasound and Mammography Images via Fusion of Deep Learning and Handcraft Features. *Entropy*, *25*(7). https://doi.org/10.3390/e25070991

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Jabeen, K., Khan, M. A., Hamza, A., Albarakati, H. M., Alsenan, S., Tariq, U., & Ofori, I. (2024). An EfficientNet integrated ResNet deep network and explainable AI for breast lesion classification from ultrasound images. *CAAI Transactions on Intelligence Technology*. https://doi.org/10.1049/cit2.12385

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data--recommendations for the use of performance metrics. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 245–251. https://doi.org/10.1109/ACII.2013.47

Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13064006

Krithiga, R., & Geetha, P. (2021). Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, *28*(4), 2607–2619. https://doi.org/10.1007/s11831-020-09470-w

Luong, H. H., Nguyen, H. T., & Thai-Nghe, N. (2024). AEGANB3: An Efficient Framework with Self-attention Mechanism and Deep Convolutional Generative Adversarial Network for Breast Cancer Classification. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Issue 5). DOI: 10.14569/IJACSA.2024.01505139.

Ma, D., Wang, C., Li, J., Hao, X., Zhu, Y., Gao, Z., Liu, C., Luo, C., & Huang, Y. (2025). Analysis of the diagnostic efficacy of ultrasound, MRI, and combined examination in benign and malignant breast tumors. *Frontiers in Oncology*, *15*, 1494862. https://doi.org/10.3389/fonc.2025.1494862

Mersha, M., Lam, K., Wood, J., AlShami, A., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing*, 128111. https://doi.org/10.1016/j.neucom.2024.128111

Nasir, M. U., Ghazal, T. M., Khan, M. A., Zubair, M., Rahman, A. U., Ahmed, R., Hamadi, H. Al, & Yeun, C. Y. (2022). Breast Cancer Prediction Empowered with Fine-Tuning. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/5918686

Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2003). Region of interest based analysis of functional imaging data. *Neuroimage*, *19*(4), 1303–1316. https://doi.org/10.1016/S1053-8119(03)00188-5

Pathan, R. K., Alam, F. I., Yasmin, S., Hamd, Z. Y., Aljuaid, H., Khandaker, M. U., & Lau, S. L. (2022). Breast Cancer Classification by Using Multi-Headed Convolutional Neural Network Modeling. *Healthcare (Switzerland)*, *10*(12). https://doi.org/10.3390/healthcare10122367

Podda, A. S., Balia, R., Barra, S., Carta, S., Fenu, G., & Piano, L. (2022). Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. *Journal of Computational Science*, *63*. https://doi.org/10.1016/j.jocs.2022.101816

Rajaan, R., Kumar, L., Soni, K., & Butwall, M. (2024). Anomaly Detection in Medical Images Using SMOTE Algorithm: A Comprehensive Approach. *International Journal of Electrical, Electronics and Computers*, *9*(5), 15–19. https://doi.org/10.22161/eec.95.2

Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., & Qadir, J. (2022). Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Computers in Biology and Medicine*, *149*, 106043. https://doi.org/10.1016/j.compbiomed.2022.106043.

Sahu, A., Das, P. K., & Meher, S. (2024). An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomedical Signal Processing and Control*, *87*. https://doi.org/10.1016/j.bspc.2023.105377.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. *International Conference on Machine Learning*, 10096–10106. https://doi.org/10.48550/arXiv.2104.00298

Uzair Khan. (2021). Breast cancer patients MRI's. Kaggle. https://www.kaggle.com/uzairkhan45/breast-cancer-patients-mris.

Vigil, N., Barry, M., Amini, A., Akhloufi, M., Maldague, X. P. V., Ma, L., Ren, L., & Yousefi, B. (2022). Dual-Intended Deep Learning Model for Breast Cancer Diagnosis in Ultrasound Imaging. *Cancers*, *14*(11). https://doi.org/10.3390/cancers14112663

Waks, A. G., & Winer, E. P. (2019). Breast cancer treatment: a review. *Jama*, *321*(3), 288–300. https://doi.org/10.1001/jama.2018.19323

Wang, J. Z. (2001). Integrated Region-Based Image *Retrieval*. Springer US. https://books.google.iq/books?id=pG1nzW9dYz4C. https://doi.org/10.1007/978-1-4615-1641-5

Zhang, J., Zhang, Z., Liu, H., & Xu, S. (2023). SaTransformer: Semantic-aware transformer for breast cancer classification and segmentation. *IET Image Processing*, *17*(13), 3789–3800. https://doi.org/10.1049/ipr2.12897