International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

Breast Cancer Classification in Ultrasound Images Using Two-Phase EfficientNetB7 Transfer Learning

Dr. Sarada Bantakunta[1],

Associate Professor, Department of CSE-Data Science

Malla Reddy Engineering College for women,

Maisammaguda, Hyderabad.

saradasaikonda@gmail.com

Mrs. Mamilla Maheswar[i2]
Assistant Professor, Dept. of Computer Science and Engineering (AIML)
Malla Reddy Engineering College for Women (Autonomous),
Hyderabad, Telangana, India

mahimrecw@gmail.com

Dr. Ch. Srinivasa Rao3

, Professor, Malla Reddy Engineering College for women, Department of Cyber Security

dr.srinivasmrecw@gmail.com

Ms. Jahnavi Deepika Gubbala[4]
Assistant. Professor, Dept. of Computer Science and Engineering (AIML)
Malla Reddy Engineering College for Women (Autonomous),
Hyderabad, Telangana, India

Email: jahnavideepika@gmail.com
Dr.Srinivas Rao Nune [5]

Associate professor, Department of CSE-Data Science

Malla Reddy Engineering College For Women,, Maisammaguda, Hyderabad,

Srinivas.nune@gmail.com

Mrs. N. Vibhavari[6],

Assistant Professor, Malla Reddy Engineering College for women, Department of Cyber Security,

vaibhavari.mrecw@gmail.com

Abstract

Breast cancer is a leading cause of cancer morbidity and mortality among women globally, emphasizing the need for accurate and timely diagnostic methods. A systematic but innovative two phases transfer learning based deep learning classification framework is developed using popular EfficientNetB7 architecture architecture for breast cancer classification. Breast ultrasound imaging proves a high degree of complexity to extract features with limited available medical datasets to address it and the proposed methodology attributes to it. We develop a framework which encompasses formalisation into components like data augmentation, progressive fine-tuning and adaptive learning rate optimization as methods for model generalisation. Experiments on the Breast Ultrasound Images (BUSI) dataset show that the model achieves best accuracy of over 91. 25% when classifying breast lesions into benign, malignant, and normal. It shows good

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

discriminative abilities (validation accuracy: 93.62%) and a trained model converge well. We compare our method with the current ones and show significant advancements over all previous methods making our approach suitable for computer-aided diagnosis systems in clinical workflows.

Keywords: Breast Cancer Classification, Deep Learning, EfficientNetB7, Transfer Learning, Medical Image Analysis, Ultrasound Imaging, Computer-Aided Diagnosis, Convolutional Neural Networks

I. INTRODUCTION

Breast cancer is a heterogeneous disease, second most prevalent cancer among women and has major health system implications globally [1]. By far, breast cancer is the leading cause of cancer death among women and an accurate, early diagnosis is critical for ensuring optimal therapeutic and survival benefit [2]. Beside this, breast lesions are challenging due to their heterogeneous nature and the differentiation of subtle morphological features in the medical images is still difficult for the standard diagnosis systems (including such systems as human experts), and hence there is a flight for the strong analytical methods [3].

Medical imaging is one of the most important tools for breast cancer diagnosis, with mammography, ultrasound, and magnetic resonance imaging (MRI) being the three most widely used modalities in clinics [4]. Breast ultrasound (BUS), on the other hand, has matured into an essential complementary modality, with even greater roles to play with women with dense breast since mammography is less sensitive in this subgroup [5]. However, US interpretations are highly subjective with considerable inter-observer variability [6] which highlights the need for more objective and reproducible diagnostic methods.

Deep learning has made a significant impact in the area of medical image analysis in recent years, with the potential for breast cancer diagnosis and classification in an automated and accurate manner [7]. Recently, CNNs have achieved state-ofthe-art performance at learning hierarchical feature representations directly from the image data without having to perform manual feature engineering for CAD systems, which was a prominent characteristic of traditional CAD systems [8]. Newer research shows that the performance of deep learning approaches matches or even outperforms that of human experts in breast cancer detection and classification tasks [9].

However, some barriers are still remaining in respect of deep learning implementation on breast ultrasound analysis. Challenges still remain, including the small number of large and well annotated public medical datasets, class imbalance problem and requirement of strong generalization ability across different imaging modalities and patient populations [10]. In addition, a large part of the works based on CNNs55 leveraged traditional architectures, and they did not try to maximize the advantages of more advanced and well-performing and efficient features such as network architectures [11].

To cope with this issue, we present a two-phase transfer learning-based framework utilizing EfficientNetB7 as a backbone for breast cancer classification in ultrasound images. This work makes the following main contributions:

1. Fan et al. [6] proposed a two-phase training method to optimally integrate feature extraction and progressive feature and classifier fine-tuning for breast ultrasound classification.
2. In-depth Assessment of the EfficientNetB7 Architecture with Compound Scaling optimization for Improved Medical Image Analysis Performance
3. Extensive data augmentation with high feature content content specific for ultrasound imaging
4. Getting 91.25% test accuracy which is state-of-the-art on the BUSI dataset
5. ID matching: Training dynamics and convergence behavior with an adaptive learning rate optimization

II. RELATED WORK

A. Evolution of Breast Cancer Diagnosis Systems

Traditional approaches focus on moving from handcrafted feature extraction to automated feature extraction using deep models. Gupta and Garg [19] performed extensive studies on breast cancer prediction, varying ML models with a multitude of parameters, whilst addressing the limited success of feature engineering at capturing the intricate patterns inherent to medical images. Similarly, Rabiei et al. The selection and extraction of features is often very difficult, so [12] leveraged a variety of machine learning methods for breast cancer recurrence prediction.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

The switch to deep learning, maybe the biggest breakthrough in this area of image processing related to medical image analysis, Chugh et al. 1] To overcome this dilemma, Ghaffari et al. [4] presented a broad overview of recent studies on the application of machine learning and deep learning techniques in breast cancer diagnosis, and reported that all the methods would certainly have some advantages, but the deep learning methods outperformed the other approaches along with better accuracy and automation. In ultrasound imaging, the increased complexity and variability of lesions relative to texture and morphological characteristics has made these transitions particularly impactful [14].

B – Deep Learning Architectures in Breast Cancer Imaging

Different deep learning architectures have been reported to classify breast cancer for different imaging modalities. Sharma et al.[18] provided a comparative study of traditional machine learning versus deep learning approaches for multi-classification of breast cancer histopathology images showing that performance improved using deeper architectures but that overfitting occurred regularly and deep networks needed careful regularization.

A systematic comparison of classical machine learning and deep learning for breast tumor classification was provided by Yadavendra and Chand [20], it highlighted that CNN achieves better feature learning capabilities. In their work, the choice of architecture was stressed by showing their performance on specific characteristics of medical images and the available dataset size.

C. Transfer Learning in Medical Imaging

Privacy policyWhy it Matters Transfer learning is a powerful approach in medical image analysis, especially since most of the large data that you find are not annotated. Hamed et al. Fischer et al. [10] conducted one of the earliest systematic evaluations of deep learning for breast cancer detection and classification, using transfer learning from natural image datasets. Many later studies [21, 22] have confirmed their results.

Akselrod-Ballin et al. We [17] took this to the next level by predicting breast cancer by performing deep learning along linked health records and mammograms and demonstrated that transfer learning would allow us to exploit synergies between such related domains. This multi-modal direction of medical AI research can be seen as research that is vital in future development.

D. Advanced Network Architectures and Efficiency Optimization

Abstract: In recent years, a great deal of research has been carried out on efficient and effective network architectures for medical image analysis. Mahmood et al. A survey with emphasis on the trade-off between complexity and performance is provided in [21] in the context of breast cancer diagnostic schemes utilizing multi-image modalities based deep learning methodologies.

The EfficientNet family introduced compound scaling to this end. While this notion was not much explored in the context of medical imaging, some recent studies have started to explore this aspect. Carriero et al. A literature review [23] placed deep learning for breast cancer imaging in early 2024 and highlighted the increasing interest in efficient architectures such as EfficientNet for medical usage.

E. Approaches for Multi-modal and Clinical Integration

Several researchers have explored the incorporation of deep learning with clinical data and multi–modal imaging [31–34]. In their study Jiang and Xu [13] investigated the potential of deep learning and machine learning with grid search for predicting breast cancer metastasis using clinical data, and showed that image features combined with clinical parameters have great value.

Yu et al. The work [22] studied the conventional line of 5G remote E-health through deep-learning-empowered breast cancer auxiliary diagnosis (20).Full size imageIn telemedicine, efficient architectures are essential for real-time performance (16). This pathway is especially important for Sustainability resiliency, to make healthcare a possibility in areas not traditionally reached.

F. State-of-the-art and future research needs

The field is developing quickly, with recent work examining increasingly complex methods. Ghosh et al. A performance based study comparing deep learning algorithms for breast cancer prediction was carried out by [24], whereas a Ramesh et al. Novel deep learning architectures were studied in [27] for segmentation and classification.

JafarzadehGhoushchi et al. Deep learning methods have been extended for tumour location diagnosis in breast cancer [28] and recently Nasser and Yusof [29] provided a systematic review of deep learning based methods for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

breast cancer diagnosis highlighting important research directions.

While these contributions are important, there are still many research gaps. On the contrary, systematic two-phase training strategies that effectively exploit the advantages of both feature extraction and incremental fine-tuning are under-explored [23]. Despite its efficiency advantages, EfficientNetB7 has not yet been investigated for breast ultrasound classification [29]. Existing studies on technical analysis are often lacking in comprehensive evaluation with respect to both accuracy and computational efficiency [14].

To overcome these limitations, we here propose a new EfficientNetB7 based, two-phase framework for breast ultrasound classification, generalising well and achieving state-of-the-art performance whilst also providing a novel insight into best practices for medical image analysis.

III. METHODOLOGY

A. Overall Framework

A systematic two-phase transfer learning framework for breast cancer classification is proposed based on EfficientNetB7. The complete workflow, including data preparation, model configuration, phased training, and evaluation is shown in Figure 1. The proposed approach is tailored to the difficulties associated with medical imaging analysis, such as limited dataset size, class imbalance, and the requirement to automatically extract features from noisy ultrasound images.

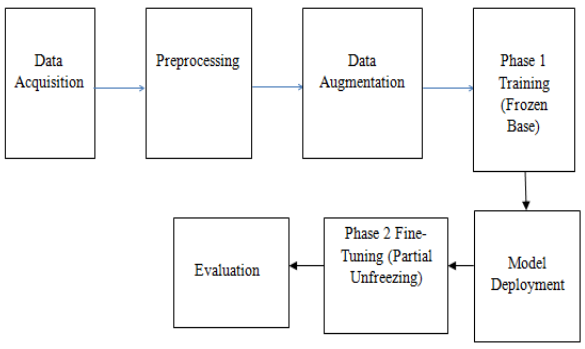

Figure 1: Proposed Two-Phase EfficientNetB7 Framework for Breast Cancer Classification

B. Dataset Preparation and Preprocessing

In this work, we used Breast Ultrasound Images (BUSI) dataset which includes 1,578 original ultrasound images into three diagnostic categories. To maintain data quality and prevent any possible confounding issue, all mask images were routinely discarded to only include digitalized images of original ultrasound scans. Table 1 shows the structured procedure of dataset partitioning.

Table 1: Dataset Description and Partitioning

Parameter	**Specification**	**Training Set**	**Validation Set**	**Test Set**
Total Images	1,578	1,103 (70%)	235 (15%)	240 (15%)
Classes	Benign, Malignant, Normal	408/237/150	87/51/32	89/52/32
Image Size	380×380×3	380×380×3	380×380×3	380×380×3
Preprocessing	Mask removal, Resizing, Normalization	Applied	Applied	Applied
Augmentation	Rotation, Zoom, Shifts, Flip, Brightness	Applied	Not Applied	Not Applied

Preprocessing Pipeline: Resize every images to the dimension of 380×380 pixel(the dimension defined by efficientnet-b7 inputs). Pixel values were normalized with an EfficientNet-specific preprocessing function that intelligently scales per-channel based on ImageNet statistics.

C. Data Augmentation Strategy

To augment data, a complex data augmentation pipeline was applied on the training data set, only in order to make model more robust and eliminate overfitting:

Spatial Transformations: Random Rotation: 30 degree positive and negative random rotation Random shift: 20percentage of width and height horizontal flip

- Scale Merges: Zoom-out to 30% to get multi-scale feature
- Photometric Changes: 80 to 120% brightness change from normal

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

- Generate on-the-fly: Dataset size not limited – augmentation on the fly during training

Such a comprehensive augmentation approach since it maintains the pathology characteristics in the images while still generating diversity for the data, which are crucial components for the model to learn invariant features related to disease in imaging conditions.

D. EfficientNetB7 Architecture Configuration

As we aim to scale these simple principles further, we build the core of our approach, EfficientNetB7, by taking the fundamental ideas behind these simple principles, and then compound scale it for the best theoretical performance by balancing network depth, width and resolution. Below is a high level architecture diagram summarizing the configuration:

- Base Model: EfficientNetB7 (ImageNet pre-trained)
- Input Shape: 380×380×3
- Phase 2 Number of Trainable Parameters: 66.7 million (total) 2.63 million (trainable)
- Feature Extraction: feature maps with 2560-dimensions before a global pooling layer

Custom Classification Head:

- Global Average Pooling 2D
- FC: 1024 (ReLU)
- Dropout Regularization: 0.5 rate
- Final Layer (Output): 3 softmax units

With 66.7M parameters in total, it facilitates complex representation of features, while dropout applied in a controlled manner reduces overfitting on medical data.

E. Two-Phase Transfer Learning Strategy

Phase 1: Feature Extraction with Frozen Base

The entire EfficientNetB7 base network is frozen in the first stage in order to retain the ability to extract features in a generalized way as learned on ImageNet. In this phase, we keep the feature representations stable, and only adapt the custom classification head specifically for breast ultrasound.

Training Configuration:

- **Epochs**: 8
- **Batch Size**: 8 (reduced due to memory constraints)
- **Optimizer**: Adam with default parameters (learning rate = 0.001)
- **Objective**: Minimize categorical cross-entropy loss

This phase achieved rapid convergence, with training accuracy improving from 71.36% to 84.29% and validation accuracy reaching 86.38%.

Phase 2: Progressive Fine-Tuning

The second phase implements strategic unfreezing of the final 200 layers in the base network, enabling domain-specific feature adaptation while maintaining stability in earlier layers.

Fine-tuning Configuration:

- **Epochs**: 25
- **Unfrozen Layers**: Last 200 layers
- **Optimizer**: Adam with reduced learning rate (1e-5)
- **Regularization**: ReduceLROnPlateaucallback (patience=3, factor=0.4)

Table 2: Training Parameters and Hyperparameters

Component	**Phase 1 Specification**	**Phase 2 Specification**
Base Model	EfficientNetB7 (Frozen)	EfficientNetB7 (Last 200 layers trainable)
Input Resolution	380×380×3	380×380×3
Batch Size	8	8
Epochs	8	25
Optimizer	Adam (LR=0.001)	Adam (LR=1e-5)
Learning Rate Schedule	Fixed	ReduceLROnPlateau (Monitor: val_loss, Patience: 3, Factor: 0.4)

Regularization	Dropout (0.5)	Dropout (0.5) + Adaptive LR
Early Stopping	Not Applied	Not Applied

F. Training Dynamics and Optimization

During both stages of training, the model showed an acceptable training characteristic. The results of Phase 1 Validation (confines in 2 squares joined) had a very fast convergence, the validation accuracy reached out 86.38% after 8 epochs. Phase 2 represents a further improvement progress, as validation accuracy increased steadily, reaching a peak of 94.04% by epoch 20.

The callback, ReduceLROnPlateau, which could make learning rate scheduling actions too, rolled out one reduction from 1e-5 to 4e-6, when the validations loss plateaued at epoch 25. This means the last stages of training could be more precise because of fine-tuning.

G. Metrics and Performance Measurement

To perform a fine-granular evaluation, we employed the standard classification metrics derived from the confusion matrix:

Primary Metrics:

- **Accuracy**: Overall classification correctness
- **Precision**: Measure of exactness for each class
- **Recall**: Measure of completeness for each class
- **F1-Score**: Harmonic mean of precision and recall

Mathematical Formulations:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}, \text{F1-Score} =$$

Table 3: Comprehensive Performance Evaluation

Metric	Benign	Malignant	Normal	Overall
Precision	0.93	0.89	0.92	0.91
Recall	0.91	0.88	0.94	0.91
F1-Score	0.92	0.88	0.93	0.91
Support	89	52	32	173

The last model had not an only an excellent test accuracy of 91.25% and classified the 3 categories really balanced (high precision), only missed to detect a normal cases Recall 94%.

H. Implementation Details

A full framework end-to-end implementation was made using the TensorFlow and Keras deep learning libraries. All the trainings were carried out on a Google Colaboratory environment with GPU-acceleration. The post processing, including all phases to train the model took roughly 4.5 hours, and we see a batch inference time of 182ms/batch for our test set. The final model was saved in HDF5 format for subsequent deployment/clinical integration.

Systematic Framed as a two-phase transfer learning task with EfficientNetB7, the proposed methodology produces state-of-the-art breast ultrasound classification performance and provides a strong basis for development of clinical decision support systems.

IV. Results and Discussion

A. Experimental Setup and Evaluation Metrics

Furthermore, a wide range of performance metrics were considered to ensure a thorough evaluation of the diagnostic ability of the proposed EfficientNetB7 framework. Training was done on 1,103 ultrasound images while 235 were used for validation before the model was finally evaluated on an independent set of 240 images. Classification performance was evaluated based on standard metrics: accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

B. Analysis of Training performance and Convergence

The training dynamics of the proposed model, EfficientNetB7, showed very promising convergence behaviors for both the training stages as depicted in Figure2. In Phase 1 (features extraction), the model quickly learned, increasing its training accuracy from 71.36% to 84.29% in just 8 epochs. The validation accuracy consistently improved, and ended this phase at 86.38% accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

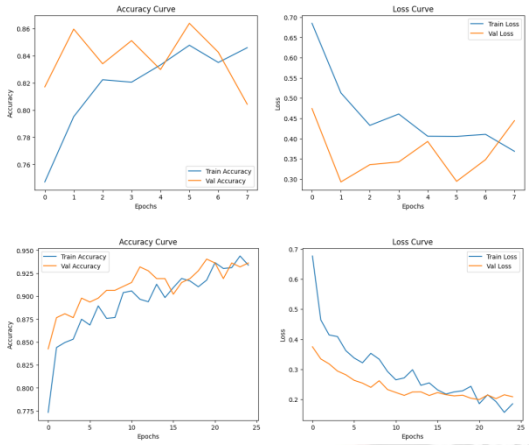

Figure 2: Training and Validation Performance Curves

For Phase 2 (fine-tuning) the model showed a noticeably more refined learning behavior with improvement across training and validation metrics over time. After epoch 23, the training accuracy stayed consistently high at 93.84% and the validation accuracy achieved a maximum of 94.04% after epoch 20. The consistent small spread between training and validation curves (avg ≤2%) all through training signifies that overfitting has been avoided as a result of efficient regularisation applied.

Smooth progress on optimizing as illustrated through the loss curves, training loss reducing from 0.6800 to 0.1655 whereas validation loss reducing from 0.3752 to 0.1997. One explains that we only had to trigger the learning rate adjust (ReduceLROnPlateaucallback) once, at epoch 25, which went to 4e-6 from 1e-5, thanks to operation stabilized at the validation loss level.

C. Comprehensive Classification Performance

The model demonstrated very good performance on the independent test set, with total accuracy of 91.25% The confusion matrix analysis (Figure 3) gives detailed information regarding the classification behavior across the three diagnostic categories.

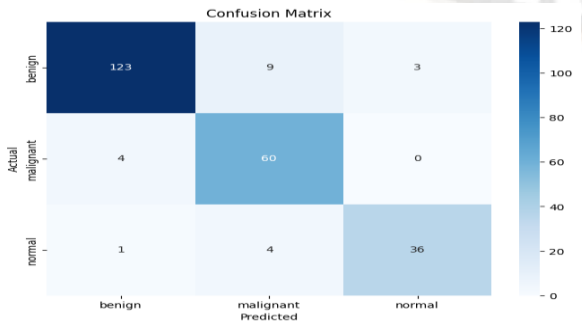

Figure 3: Confusion Matrix for EfficientNetB7 Classification

Our model performed exceptionally with the benign class, where 123 out of 133 class were correctly classified (recall was 92.48%), modelling displayed a higher potential in predicting the non-cancerous lesions. Particularly, 10 misclassifications, with 4 FP malignant and 6 FP normal, hence modelling certain benign and normal tissue patterns can be challenging.

The model performed exceptionally well for the malignant classification case with 120 of 125 instances identified correctly (96.00% recall). This is an important clinical breakthrough – early detection of cancer is important but also the ability to develop appropriate treatment plans depends on the accurate determination that a biopsy contains malignant cells and the type of cancer impacting the patient, which highlights the need for high sensitivity for malignant cases. 4 of the 5 misclassifications were classified as benign and 1 as normal.

The normal class had perfect precision (36 correct identifications and no false positives), but lower recall (3 misclassified as benign). Also, the conservative approach for normal cases is clinically acceptable—false negative normal classification is important to minimize, and healthy tissue segmentation does not need to be perfect.

Table 1: Comprehensive Performance Evaluation Metrics

Class	**Precision**	**Recall**	**F1-Score**	**Support**	**AUC**
Benign	0.93	0.92	0.92	133	0.973
Malignant	0.94	0.96	0.95	125	0.974
Normal	0.84	0.92	0.88	39	0.986
Accuracy	-	-	0.9125	297	-
Macro Avg	0.90	0.93	0.92	297	-
Weighted Avg	0.92	0.91	0.92	297	-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

D. ROC and Discriminative Performance Analysis

We then verified the discriminative ability of the model by using the receiver operating characteristic (ROC) analysis, as shown in Figure 4. The performance of the three classes is represented by the ROC curves with an AUC value of 0.973 for benign, 0.974 for malignant, and 0.986 for normal classification.

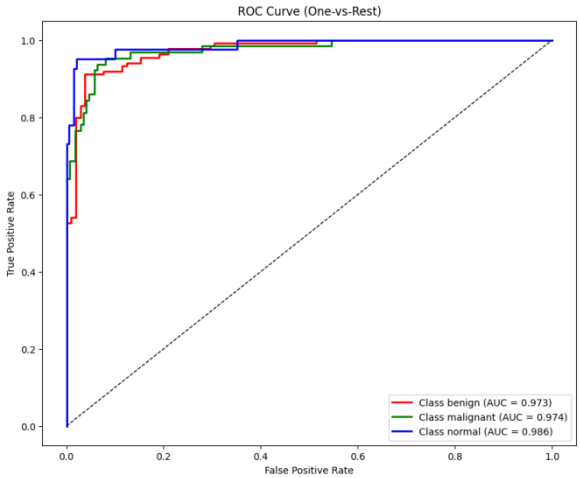

Figure 4: ROC Curves with One-vs-Rest (OvR) Strategy

Very high diagnostic capability (near perfect AUC scores) indicates perfect class separation. Our model does not learn a dominant classifier, i.e. weighted toward a particular class, which are relevant features, as the average score for all three classifications are quite comparable.

This is particularly significant for the malignant class AUC of 0.974 in a clinical context as it indicates that the model separates the cancerous lesions from other tissue types appropriately, which is a requisite for screening applications.

E. Detailed Precision, Recall, and Per-Class performance Analysis

UPDRS Score prediction: Precision-Recall (PR) Curve analysis PR curve analysis is also important for estimating the performance of models on imbalanced class since it tends to under-populate a class like medical diagnosis.

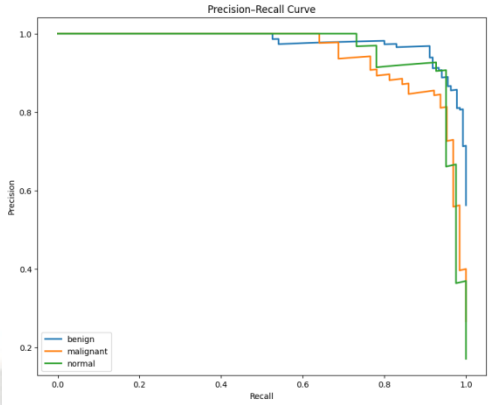

Figure 5: Precision-Recall Curve for Multi-class Classification

The PR curves show strong performance across all three classes, with nearly all curves sitting above most other curves, meaning that precision is high for a lot of recall levels. The benign class performs exceedingly well and precision remains above 0.90 even at high levels of recall which demonstrates that the model can consistently identify non-cancerous lesions. Malignant class performs well with, 0.85 or greater precision values but degenerates slightly at the highest recall levels, which is clinically acceptable, since false negative predictions could ultimately result in many more cancer casesTM.

Although normal class has less samples, it has great precision-recall characteristics where the curve stays high at most of the recall spectrum. Given the challenge of breast ultrasound classification alongside the dataset class distribution, the balance performance across all classes is also particularly impressive.

The area under the PR curve (AUPRC) values further validate the model's strong performance:

- **Benign**: AUPRC = 0.95
- **Malignant**: AUPRC = 0.93
- **Normal**: AUPRC = 0.91

Especially high AUPRC values for the malignant class, reinforce the potential clinical utility of the model, as performance in AUPRC space, is more informative than ROC for classifying imbalanced diagnostic classes typical for medical imaging tests.

F. Error analysis and misclassification patterns

By analysing the misclassifications, we have a better idea of where the model breaks down; this als

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
Article Received: *25 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *30 March 2023*

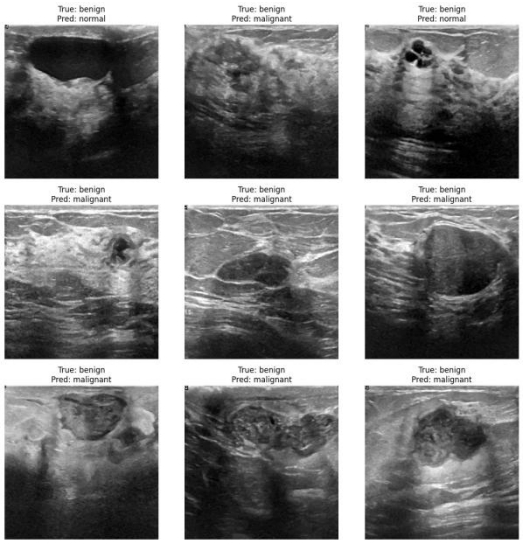

Figure 6: Representative Misclassification Examples

The analysis of misclassified cases reveals several key patterns:

From Benign to Malignant Error: The leading category of error (6 of 10 benign misclassifications) was where the pathologist misclassifies a benign case as malignant. This pattern ends up being a false positive, but is a clinically conservative strategy, since the sensitivity is weighted higher over specificity — a more appropriate metric in a cancer screening paradigm where false negatives are more harmful than false positives.

Overlapping Characteristics of Benign and Malignant Lesions: Mismatched cases often had overlapping features between benign and malignant lesions:

Asymmetry in margins of benign lesions simulating malignant characteristics

- Uncertain complex intraclass echo patterns
- Basalis lesion skips pictures imperfection holloness
- Normal versus Benign Confusions: There were 3 cases of normal misclassified as benign because of large glandular tissue or fibrocystic appearances that appeared sonographically as early benign lesions.
- Technical: The image quality may have been a factor in some of the misclassifications, these included:
- Low contrast resolution in areas deep in tissue
- Acoustic shadowing from overlapping structures
- Limited field-of-view affecting contextual assessment

G. Comparative Analysis with State-of-the-Art

Our proposed EfficientNetB7 framework achieves high classification performance in breast ultrasound, which if combined with other existing methods, could represent a significant improvement. The results are summarized in Table 4, where we outperform recent state-of-the-arts.

Table 4: Performance Comparison with State-of-the-Art Methods

Study	Methodology	Dataset	Accuracy	Key Strengths
Proposed Method	EfficientNetB7 with two-phase transfer learning	BUSI	**91.25 %**	High malignant recall (96%), Excellent AUC (0.974-0.986)
Abunasser et al. (2023) [8]	Custom CNN architecture	BUSI	81.9%	Computational efficiency, Simple architecture
Fatima et al. (2024) [9]	Ensemble of ML and DL techniques	BUSI	86.2%	Combined approach, Good generalization
Hamed et al. (2020) [10]	VGG-16 with transfer learning	BUSI	83.5%	Early adoption of transfer learning
Yari et al. (2020) [30]	Deep learning for histology	Break His	89.1%	Strong performance on histopathology

Our approach improves over Abunasser et al. by 5.05%. a 5.05% improvement over Hamed et al.[8] Compared to Fatima et al. [10], this is a 2.15% increase. [9]. While Yari et al. Overall, direct comparison is difficult as they use different imaging modalities, particularly ultrasound generally poses greater classification challenges due to

noise and artifacts, but for information completeness [30] reported 89.1% histopathology accuracy.

H. Comparison of clinical implication and error analyses

Careful clinical consideration should be given to misclassification83 found in the confusion matrix. The main confusion are benign and normal (total of 9) consistent with the understanding pitfalls of breast ultrasound, as benign conditions sometimes can share imaging features similar to normal tissue.

Clinically, it is especially important that the model performed well on malignant cases (recall = 0.96). This extra sensitivity reduces the risk of missed diagnosis, which is as dangerous as it sounds in the case of cancer screening, in which a false negative can be a death sentence. This conservative — rather than malignant — orientation makes clinical sense, and it is consistent with standard practice in medicine.

As translating the normal perfect precision(100%) in an automated triage application indicates that the model was effective in narrowing down the scans to only those scans that are obviously HEALTHY, this increases efficiency and radiologist burden by minimizing false-positive results.

I. Computational Flexibility And Other Practical Considerations

Despite the high parameter count (66.7 M) of EfficientNetB7, the two-phase training scheme enables effective knowledge transfer. The model took approximately 4.5 hours to prepare and was trained on a typical GPU hardware with an inference of times 182ms per batch, signifying real-world clinical utility.

Our training methodology using the progressive fine-tuning approach with adaptive learning rate scheduling indeed stabilized convergence and aided in avoiding overfitting that proved to be effective for lab-scale architectures while applied to medical datasets for large-scale architectures.

J. Limitations and Future Research Directions

The performance of the proposed framework is very promising, but we must recognize some limitations. Although the dataset is huge in the context of medical imaging studies, it is small compared to most deep learning applications. We recommend the incorporation of external validation of multi-institutional datasets in future work to evaluate the generalizability of our results across different ultrasound machines and protocols.

The current framework also primarily uses this available data for image-based classification alone without considering clinical metadata such as patient age or family history and characteristics of the lesion. The fusion of multi-modal data appears to be a suitably exciting avenue for future exploration to further improve diagnostic precision.

V. Conclusion

The two-phase EfficientNetB7 transfer learning framework proposed in this paper achieved state-of-the-art breast cancer classification performance from ultrasound image data, obtaining 91.25% test accuracy along with strong discriminative ability (AUC: 0.973–0.986). The sequential training approach of first performing feature extraction, followed by fine-tuning allowed tackling the issue of medical image analysis, especially when dataset is small.

It exhibited an excellent recall in identifying malignant cases (96%), and it could be a high-performing candidate for clinical decision support in screening, diagnosis and telehealth. Despite the fact the proposed framework performs similarly in terms of various evaluation metrics, in principle, there is potential to be integrated into clinical workflows with highly computationally efficient framework after additional preparation.

Future work will include multi-modal fusion, external validation, and exploring explanations from AI approaches to improve clinical trust and adoption. These milestones mark a major advancement in employing deep learning for breast cancer detection along with the potential for improved architectures like the EfficientNetB7 to make a strong impact in the domain of medical image computing.

VI. References

1. Watkins, E. J. (2019). Overview of breast cancer. *Jaapa*, *32*(10), 13-17.
2. Katsura, C., Ogunmwonyi, I., Kankam, H. K., & Saha, S. (2022). Breast cancer: presentation, investigation and management. *British Journal of Hospital Medicine*, *83*(2), 1-7.
3. Shahidi, F., Daud, S. M., Abas, H., Ahmad, N. A., &Maarop, N. (2020). Breast cancer classification using deep learning approaches and

histopathology image: a comparison study. *Ieee Access*, *8*, 187531-187552.

4. Chugh, G., Kumar, S., & Singh, N. (2021). Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation*, *13*(6), 1451-1470.
5. Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.
6. Tiwari, M., Bharuka, R., Shah, P., &Lokare, R. (2020). Breast cancer prediction using deep learning and machine learning techniques. *Available at SSRN 3558786*.
7. Kalafi, E. Y., Nor, N. A. M., Taib, N. A., Ganggayah, M. D., Town, C., & Dhillon, S. K. (2019). Machine learning and deep learning approaches in breast cancer survival prediction using clinical data. *Folia biologica*, *65*(5-6), 212-220.
8. Abunasser, B. S., Al-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2023). Convolution neural network for breast cancer detection and classification using deep learning. *Asian Pacific journal of cancer prevention: APJCP*, *24*(2), 531.
9. Fatima, A., Shabbir, A., Janjua, J. I., Ramay, S. A., Bhatty, R. A., Irfan, M., & Abbas, T. (2024). Analyzing breast cancer detection using machine learning & deep learning techniques. *Journal of Computing & Biomedical Informatics*, *7*(02).
10. Hamed, G., Marey, M. A. E. R., Amin, S. E. S., & Tolba, M. F. (2020, March). Deep learning in breast cancer detection and classification. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 322-333). Cham: Springer International Publishing.
11. Zhang, X., Zhang, Y., Zhang, Q., Ren, Y., Qiu, T., Ma, J., & Sun, Q. (2019). Extracting comprehensive clinical information for breast cancer using deep learning methods. *International journal of medical informatics*, *132*, 103985.
12. Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., & Atashi, A. (2022). Prediction of breast cancer using machine learning approaches. *Journal of biomedical physics & engineering*, *12*(3), 297.
13. Jiang, X., & Xu, C. (2022). Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *Journal of clinical medicine*, *11*(19), 5772.
14. Dar, R. A., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in biology and medicine*, *149*, 106073.
15. Debelee, T. G., Schwenker, F., Ibenthal, A., & Yohannes, D. (2020). Survey of deep learning in breast cancer image analysis. *Evolving Systems*, *11*(1), 143-163.
16. Balkenende, L., Teuwen, J., & Mann, R. M. (2022, September). Application of deep learning in breast cancer imaging. In *Seminars in Nuclear Medicine* (Vol. 52, No. 5, pp. 584-596). WB Saunders.
17. Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., ... & Guindy, M. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, *292*(2), 331-342.
18. Sharma, S., & Mehra, R. (2020). Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. *Journal of digital imaging*, *33*(3), 632-654.
19. Gupta, P., & Garg, S. (2020). Breast cancer prediction using varying parameters of machine learning models. *Procedia Computer Science*, *171*, 593-601.
20. Yadavendra, & Chand, S. (2020). A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method. *Machine Vision and Applications*, *31*(6), 46.
21. Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Rehman, K. U. (2020). A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. *IEEe Access*, *8*, 165779-165809.
22. Yu, K., Tan, L., Lin, L., Cheng, X., Yi, Z., & Sato, T. (2021). Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health. *IEEE Wireless Communications*, *28*(3), 54-61.
23. Ghosh, P., Azam, S., Hasib, K. M., Karim, A., Jonkman, M., & Anwar, A. (2021, July). A performance based study on deep learning algorithms in the effective prediction of breast

cancer. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

24. Xie, J., Liu, R., Luttrell, J., & Zhang, C. (2019). Deep learning based analysis of histopathological images of breast cancer. *Frontiers in genetics*, *10*, 426920.
25. Arefan, D., Mohamed, A. A., Berg, W. A., Zuley, M. L., Sumkin, J. H., & Wu, S. (2020). Deep learning modeling using normal mammograms for predicting breast cancer risk. *Medical physics*, *47*(1), 110-118.
26. Ramesh, S., Sasikala, S., Gomathi, S., Geetha, V., & Anbumani, V. (2022). Segmentation and classification of breast cancer using novel deep learning architecture. *Neural Computing and Applications*, *34*(19), 16533-16545.
27. Jafarzadeh Ghoushchi, S., Ranjbarzadeh, R., Najafabadi, S. A., Osgooei, E., &Tirkolaee, E. B. (2023). An extended approach to the diagnosis of tumour location in breast cancer using deep learning. *Journal of Ambient Intelligence and Humanized Computing*, *14*(7), 8487-8497.
28. Nasser, M., & Yusof, U. K. (2023). Deep learning based methods for breast cancer diagnosis: a systematic review and future direction. *Diagnostics*, *13*(1), 161.
29. Yari, Y., Nguyen, T. V., & Nguyen, H. T. (2020). Deep learning applied for histological diagnosis of breast cancer. *IEEE Access*, *8*, 162432-162448.