USR Journal of Pharmaceutical Medical & Heatlth Sciences

Abbreviated Key Title: Uni J Pha Med H Sci
Published by USR Publisher
Email: contact@usrpublisher.com

OPEN ACCESS

Research Article

Received: 09.12.2025 | Accepted: 18.01.2026 | Published: 05.02.2026 Volume-1 | Issue-1 **Jan-Feb, 2026**

Deep Transfer Learning with ResNet50 for Early Detection of Breast Cancer from Ultrasound Images

*Eugene Danwin Chedah** *(MSc, BSc)*

Department of Microbiology, Faculty of Science, Marwadi University, Rajkot, Gujarat, India

Abstract

Breast cancer is the most frequently diagnosed cancer among women worldwide, and early, accurate detection remains central to improving survival outcomes. Conventional screening modalities such as mammography and ultrasound are limited by false-positive results, reduced sensitivity in dense breast tissue, and marked inter-reader variability. Deep learning offers a potential route to more consistent, automated interpretation of breast imaging. Here we compare a baseline six-layer convolution neural network (CNN) against a ResNet50 model adapted via transfer learning for three-class classification (normal, benign, malignant) of breast ultrasound images. Using a public dataset of 1,564 ultrasound images from 600 female patients, the fine-tuned ResNet50 model reached a validation accuracy of approximately 90%, compared with approximately 65% for the baseline CNN. Loss curves for the ResNet50 model diverged between training and validation in later epochs, indicating over fitting that we attribute to the limited size of the dataset relative to model capacity. These results support transfer learning as a substantial improvement over a shallow CNN baseline for breast ultrasound classification, while underscoring that larger, multi-centre datasets and explicit regularization are needed before such models could be considered for clinical decision support.

Keywords: *Breast cancer, deep learning, transfer learning, ResNet50, convolutional neural network, ultrasound imaging*

Copyright©2022 The Author(s): This work is licensed under a Creative Commons Attribution- Non Commercial 4.0(CCBY-NC4.0) International License.

Introduction

Breast cancer accounts for approximately 25% of all cancers diagnosed in women and remains a leading cause of cancer-related mortality worldwide (World Health Organization, 2020). Prognosis is strongly tied to stage at diagnosis: early-stage disease is associated with five-year survival approaching 99%, compared to roughly 50% for late-stage disease (De Santis et al., 2019). This gap has driven sustained interest in screening technologies capable of detecting malignancy earlier and more reliably.

Mammography remains the primary screening modality, with reported sensitivity near 80% and specificity near 90% (Lehman et al., 2017), but it is limited by false positives, reduced performance in dense breast tissue, and considerable variability between readers (Elmore et al., 1994). Ultrasound is frequently used as a complementary modality, particularly for younger patients and denser tissue, though its diagnostic accuracy for small lesions is also limited when used alone (Berg et al., 2008). These constraints have motivated growing interest in machine learning, and deep learning in particular, as a means of augmenting radiologist interpretation by learning discriminative patterns directly from image data (Rajkomar et al., 2019; Litjens et al., 2017). Convolutional neural networks (CNNs) have shown encouraging results for breast lesion classification across mammography, ultrasound, and MRI, with reported accuracies in prior studies ranging from roughly 87% to 92% depending on architecture, dataset, and imaging modality (Burt et al., 2018; Zhang et al., 2019; Kim et al., 2019; Chen et al., 2019; Huang et al., 2020). Transfer learning, in which a network pretrained on a large natural-image corpus is fine-tuned on a smaller domain-specific dataset, has been proposed as a way to partially offset the scarcity of large, labeled medical imaging datasets (Song et al., 2019). Residual architectures such as ResNet50 are attractive candidates for this approach because their skip connections mitigate vanishing-gradient problems and allow effective fine-tuning even with comparatively few training examples (He et al., 2016).

In this study, we evaluate whether transfer learning with a ResNet50 backbone improves three-class classification of breast ultrasound images (normal, benign, malignant) relative to a baseline CNN trained from scratch. We report both models' performance and examine the training dynamics of the fine-tuned model, including evidence of overfitting, to assess the practical readiness of this approach and to identify the next steps required before clinical translation could reasonably be considered.

Results

Dataset characteristics

The dataset comprised 1,564 breast ultrasound images collected from 600 female patients aged 25–75 years, each with an associated ground-truth mask, at a native resolution of 500 × 500 pixels (Al-Dhabyani et al., 2020). Images were categorized into three classes: normal, benign, and malignant. All images were resized to 256 × 256 pixels and preprocessed using a watershed segmentation algorithm to extract the region of interest prior to model training. The dataset was partitioned into 1,252 images (80%) for training and 312 images (20%) for validation.

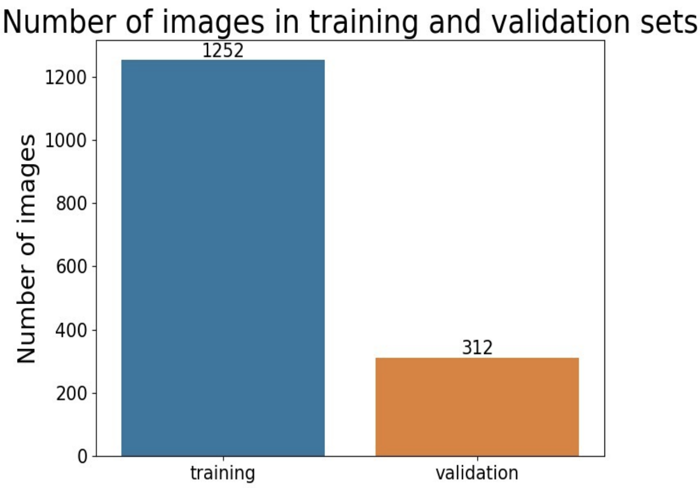

Figure 1. Number of images all

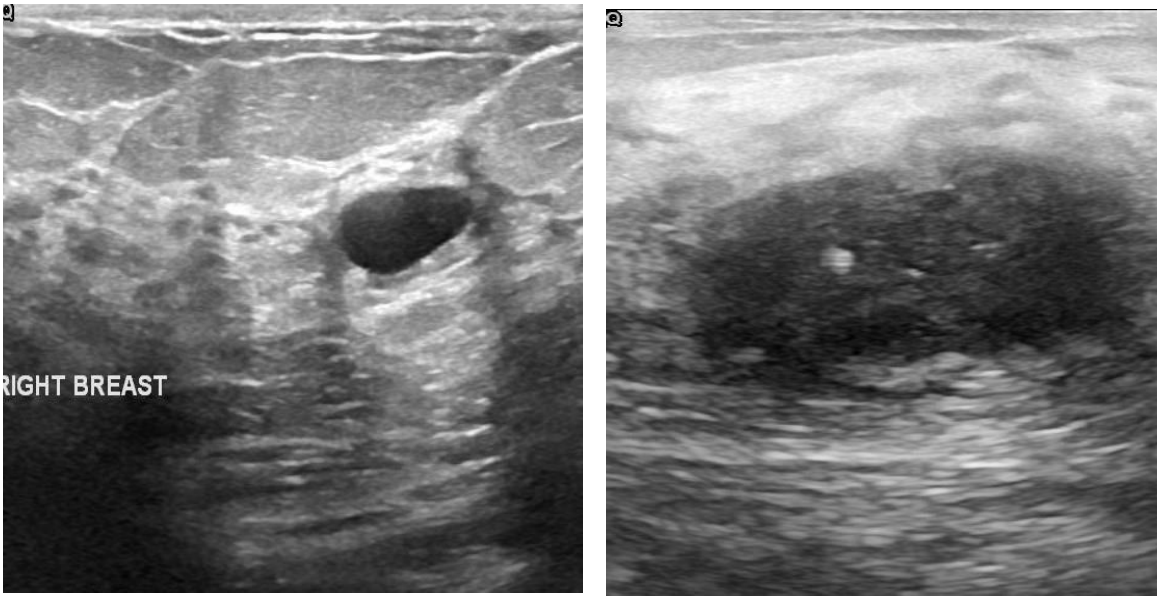

Figure 2. Representative ultrasound images from the dataset. Left: benign lesion. Right: malignant lesion.

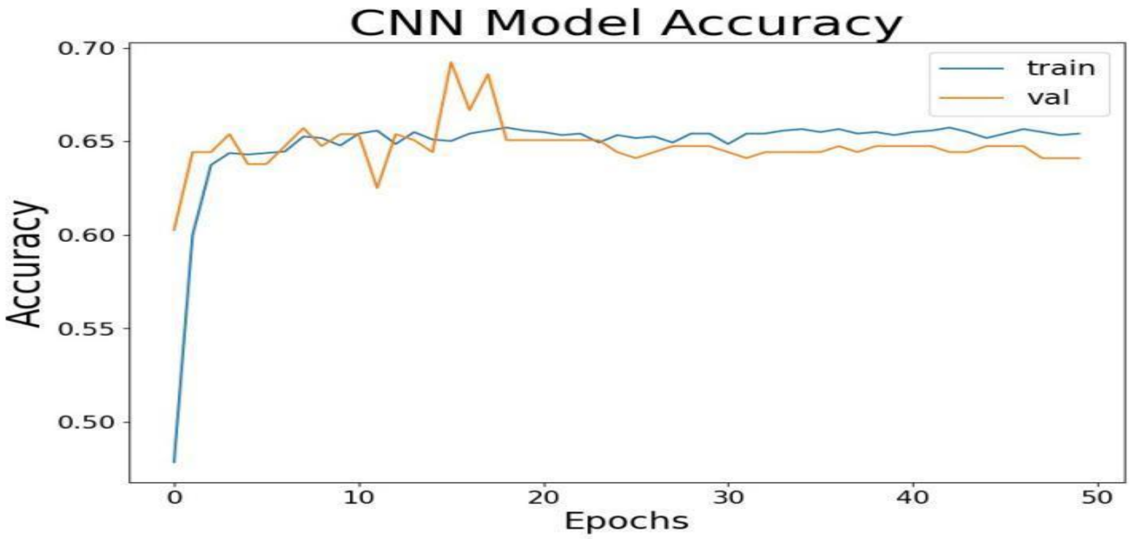

Figure 3. Representative ultrasound image of normal breast tissue.

Baseline CNN performance

A six-layer CNN trained from scratch reached a training and validation accuracy of approximately 65% after 50 epochs (Fig. 4a). Accuracy plateaued early in training, with no substantial further gain across subsequent epochs. Training and validation loss both decreased sharply within the first few epochs and remained low and closely tracked thereafter (Fig. 4b), showing no indication of overfitting at this accuracy ceiling.

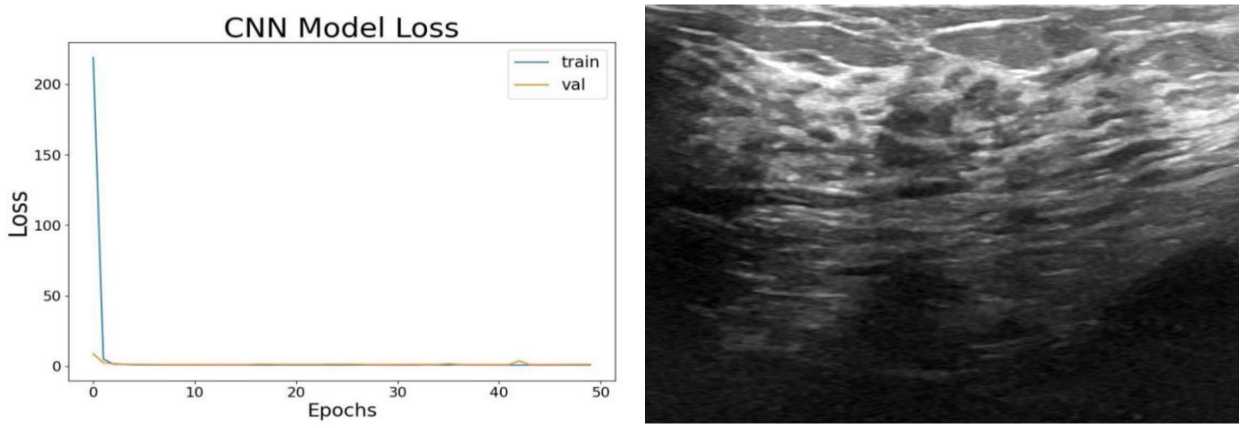

Figure 4. Baseline CNN training and validation (a) accuracy and (b) loss over 50 epochs.

ResNet50 transfer learning performance

Fine-tuning a pretrained ResNet50 model with base layers frozen and the input and output layers adapted to the three-class task substantially improved performance, reaching a validation accuracy of approximately 90% after 50 epochs, compared with 65% for the baseline CNN (Fig. 5a). However, examination of the training dynamics showed that while training loss remained low throughout, validation loss began to diverge upward and became noticeably more volatile after approximately 30 epochs, even as validation accuracy remained comparatively stable (Fig. 5b). This pattern is consistent with overfitting: the model increasingly fit patterns specific to the training set that did not generalize as reliably to held-out validation images. Given the relatively small dataset size (1,564 images across three classes), this is an expected risk for a 50-layer network of this capacity and is addressed further below.

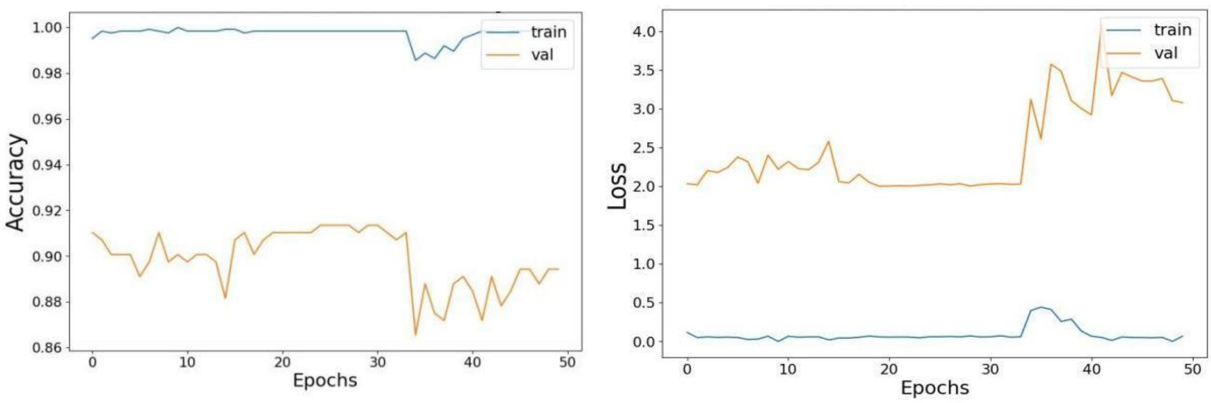

Figure 5. Fine-tuned ResNet50 training and validation (a) accuracy and (b) loss over 50 epochs. Note the divergence and increased volatility of validation loss after ~30 epochs.

Comparison of models

Across both accuracy and loss metrics, the ResNet50 model outperformed the baseline CNN throughout training (Fig. 6). The magnitude of the accuracy gap (approximately 25 percentage points) indicates that transfer learning conferred a substantial benefit over training a comparatively shallow architecture from scratch on this dataset.

At the same time, the divergence in validation loss for the ResNet50 model indicates that the reported 90% validation accuracy should be interpreted as an upper-bound estimate under current conditions rather than a stable, generalizable figure.

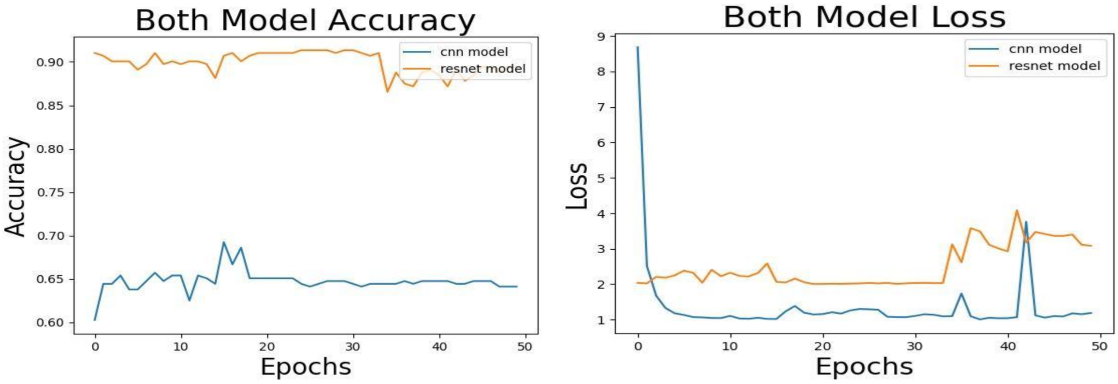

Figure 6. Direct comparison of the baseline CNN and fine-tuned ResNet50 model: (a) validation accuracy and (b) validation loss over 50 epochs.

Discussion

This study demonstrates that transfer learning with a ResNet50 backbone can substantially improve automated classification of breast ultrasound images relative to a shallow CNN trained from scratch, consistent with prior reports that transfer learning improves breast lesion classification performance across imaging modalities (Burt et al., 2018; Song et al., 2019). The accuracy achieved here (approximately 90%) is broadly in line with previously reported ranges for deep learning-based breast ultrasound classification (Zhang et al., 2019; Kim et al., 2019).

The overfitting observed in the ResNet50 model's validation loss curve is an important limitation of the present work. It most plausibly reflects the mismatch between a 50-layer network's representational capacity and a training set of just over 1,200 images spread across three classes. This is a common challenge in medical imaging applications, where labeled data is inherently costly to obtain, and it should temper direct clinical interpretation of the reported accuracy figure. We did not apply explicit regularization (such as dropout or weight decay beyond the frozen base layers) or k-fold cross-validation in the reported experiments, and we consider both to be necessary next steps rather than optional refinements.

Several additional limitations warrant note. First, the dataset, while suitable for a proof-of-concept comparison, is derived from a single source and has not been externally validated on images from other institutions or ultrasound equipment, which limits confidence in generalizability. Second, this study used ultrasound images exclusively; combining imaging modalities, as shown in prior work to improve accuracy (Chen et al., 2019), was outside the present scope. Third, model interpretability was not assessed; techniques such as class activation mapping would help clarify whether the network is attending to clinically relevant regions rather than incidental artifacts, which is particularly important given the overfitting observed.

Future work should prioritize (i) k-fold or leave-one-subject-out cross-validation to obtain more robust performance estimates, (ii) explicit regularization strategies (dropout, weight decay, early stopping tied to validation loss rather than accuracy) to address the overfitting identified here, (iii) evaluation on external, multi-center datasets, and (iv) interpretability analysis to support clinical trust. Data augmentation approaches, including generative adversarial network-based augmentation, may also help address the underlying data scarcity, though this was not implemented in the present experiments and should be evaluated directly rather than assumed to resolve the overfitting observed.

Taken together, these findings support transfer learning as a promising direction for automated breast ultrasound classification while indicating that the current model is not yet ready for clinical decision support without further validation and methodological refinement.

Methods

Dataset and preprocessing

This study used a publicly available breast ultrasound image dataset comprising 1,564 images from 600 female patients aged 25 to 75 years, each labeled as normal, benign, or malignant and accompanied by a ground-truth segmentation mask (Al-Dhabyani et al., 2020). Images were resized to 256 × 256 pixels. Prior to feature extraction, images were preprocessed using a watershed segmentation algorithm to perform region-of-interest extraction and region enhancement. The dataset was split into training (80%, n = 1,252) and validation (20%, n = 312) subsets.

Model architectures

Two models were trained and compared. The baseline model was a six-layer CNN trained from scratch on the preprocessed images. The comparison model used ResNet50 (He et al., 2016), a 50-layer residual network pretrained on natural images, adapted via transfer learning: base convolutional layers were frozen to retain pretrained feature representations, and the input and final (classification) layers were modified for the three-class classification task.

Training and evaluation

Both models were trained for 50 epochs. Training and validation accuracy and loss were recorded at each epoch to assess learning dynamics and generalization. No cross-validation, data augmentation, or explicit regularization beyond the frozen-layer transfer learning setup was applied in the reported experiments; this is noted as a limitation in the Discussion.

Data Availability

The breast ultrasound image dataset analyzed in this study is publicly available (Al-Dhabyani et al., 2020). Model training code is available from the corresponding author upon reasonable request.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H. & Fahmy, A. Dataset of breast ultrasound images. Data Brief 28, 104863 (2020).
2. Berg, W. A. et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 299, 2151–2163 (2008).
3. Burt, J. R. et al. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. Br. J. Radiol. 91, 20170545 (2018).
4. Chen, X. et al. Multi-modal deep learning for breast tumor classification combining mammography and MRI. Comput. Biol. Med. 115, 103498 (2019).
5. De Santis, C. E. et al. Breast cancer statistics, 2019. CA Cancer J. Clin. 69, 438–451 (2019).
6. Elmore, J. G. et al. Variability in radiologists' interpretations of mammograms. N. Engl. J. Med. 331, 1493–1499 (1994).
7. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In Proc. IEEE Conf. Computer Vision and Pattern Recognition 770–778 (2016).
8. Huang, Q. et al. Ultrasound imaging based on the combination of deep learning and decision tree classification for breast tumor diagnosis. Comput. Math. Methods Med. 2020, 1–9 (2020).
9. Kim, J. et al. Deep learning-based classification of breast masses on ultrasound imaging. Ultrasonography 38, 130–137 (2019).
10. Lehman, C. D. et al. National performance benchmarks for modern screening digital mammography. Radiology 283, 49–58 (2017).
11. Litjens, G. et al. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88 (2017).
12. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. N. Engl. J. Med. 380, 1347–1358 (2019).
13. Song, D. et al. Transfer learning for breast cancer detection in mammography. Comput. Methods Programs Biomed. 179, 104953 (2019).
14. World Health Organization. Breast cancer fact sheet. (WHO, 2020).
15. Zhang, Q. et al. A benchmarking study of deep learning-based breast cancer detection using mammography images. J. X-Ray Sci. Technol. 27, 791–805 (2019).