Breast Cancer Detection using Hybrid Models and Transfer Learning

Froduard Habimana
Handong Global University
558 Handong-ro, Buk-gu, pohang, Gyeongbuk 37554
hfroduard@handong.ac.kr

Djika Asmaou Houma
Handong Global University
558 Handong-ro, Buk-gu, pohang, Gyeongbuk 37554
asmaoulandi@handong.ac.kr

Jaeeun Lee
Handong Global University
558 Handong-ro, Buk-gu, pohang, Gyeongbuk 37554
Jaeeunlee.start@gmail.com

Manzi Dave Rugari
Taylor University
1864 S Main St, Upland, IN 46989
dave_rugari@taylor.edu

Abstract

Breast cancer remains a significant global health challenge, particularly for women aged 50 and older, yet it can also impact younger women and men. The accurate detection and classification of various breast cancer types, such as invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS), are critical for effective treatment and improved patient outcomes. Despite advancements in deep learning methods, including Convolutional Neural Networks (CNNs) and U-Net architectures, existing models still face challenges related to data quality, scarcity, and the urgency of timely diagnosis. This research aims to enhance breast cancer detection rates by developing optimized deep learning frameworks that leverage both CNNs and Vision transformer models. A dual analysis was done with the CBIS-DDSM and the IDC (histopathology) datasets. Data augmentation techniques like rotation, vertical flip, and zoom were also used to improve performance. The results were evaluated using existing methods on both datasets, with the best results with Xception-ViT for accuracy (96.8%), precision (96%), recall (97%), and F1-score (96%), for the CBIS-DDSM dataset and pre-trained model Prov-GigaPath with accuracy (93.23%), precision (90.42%), recall (85.15%), and F1-score (80%) for IDC dataset.

Keywords: Breast Cancer; ViT; CNN; Hybrid Model.

1. Introduction

Breast cancer was the most diagnosed cancer among women globally in 2022, with 2.3 million new cases and 670,000 deaths (WHO, 2022). Early detection can significantly improve survival rates, with up to 99% for early-stage diagnoses according to National Breast Cancer Foundation. Projections for 2040 suggest over 3 million annual cases and 1 million deaths (International Agency for Research on Cancer, 2022). Invasive ductal carcinoma (IDC) is the most common subtype, accounting for 80% of cases (Bolhasani et al., 2020). This research aims to improve breast cancer detection by experimenting with the CBIS-DDSM and IDC datasets using hybrid models and transfer learning approach to optimize accuracy. We employed various augmentation techniques to enhance model generalizability and evaluated performance using standard metrics, including accuracy, precision, recall, F1 score, and confusion matrix and at the end we provided key insights and recommendations for future studies.

2. Methodology

2.1 Dataset description

Two datasets were used in this research: the Breast Histopathology Image dataset (IDC) and The CBIS-DDSM dataset. The IDC dataset consisted of 162 whole-mount slide images of Breast Cancer (BCa) specimens scanned at 40x. From that, 277,524 size 50 x 50 patches were extracted (198,738 IDC negative and 78,786 IDC positive). The CBIS-DDSM dataset is a standardized collection of mammogram images derived from the original DDSM dataset. It contains 6,775 studies and 10,239 images, with 1,566 distinct participants. The images are stored in JPEG format and span a variety of cases, including normal, benign, and malignant breast conditions, with verified pathology information. The dataset also includes segmentation maps, bounding boxes, and annotations for regions of interest (ROIs). This curated subset of DDSM trains, tests, and evaluates machine learning models for breast cancer detection, providing a well-organized, publicly available mammography research dataset.

2.2 Preprocessing

Data preprocessing is essential for improving the quality of input images and ensuring that they are suitable for deep-learning models. Preprocessing steps include resizing the images, normalization (scaling pixel values), and augmenting the data through transformations such as rotations and flips to increase the dataset's size and improve model generalization.

2.3 Proposed framework

The updated proposed framework integrates CNNs and hybrid models (e.g., Xception+ViT, EfficientNetB0+ViT, and ProvGigaPath) to classify medical images from CBIS-DDSM and IDC datasets as benign or malignant. It includes preprocessing, data augmentation (e.g., flipping, rotation), and data splitting for training, validation, and testing. Performance is evaluated using accuracy, precision, recall, F1 score, and AUC.

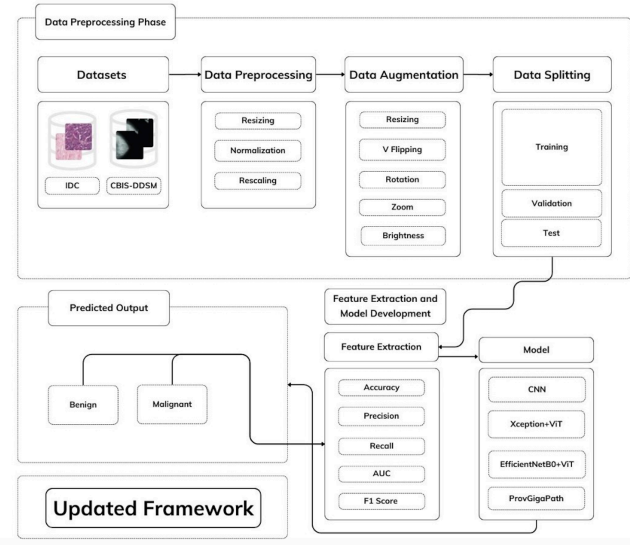

Figure 1 : Proposed framework for classification BC based on deep learning architecture

2.3.1 CNN

Convolutional Neural Networks (CNNs) form the core of the deep learning architectures applied to breast cancer detection. These CNNs consist of several layers, including convolutional layers for feature extraction, pooling layers for downsampling, and fully connected layers for classification. The most common architectures used include VGGNet, ResNet, and DenseNet, which are pre-trained on large datasets like ImageNet.

2.3.2 Xception

The Xception model, short for "Extreme Inception", is a deep convolutional neural network architecture from 2017 designed by François Chollet. It is based on Inception architecture but uses a much simpler and more efficient mechanism "depthwise separable convolutions" instead of hand-crafted modules. In a depth separable convolution, one splits a standard convolution into 2 separate convolutions: a depthwise convolution (which filters each input channel independently) followed by a pointwise convolution (which combines these filtered outputs with 1 x 1 convolutions). (Boesch et al., 2016) It consists of three parts: an entry flow that downsamples the input, a middle flow that extracts features using multiple repeated separable convolution blocks, and an exit flow that converts the learned representations to final outputs. This architecture makes image classification, feature extraction, and transfer learning tasks very efficient and accurate.

2.3.3 EfficientNetB0

EfficientNetB0 is one of the models in the EfficientNet family, which scales model size (depth, width, and resolution) with a compound scaling method to achieve state-of-the-art accuracy levels while being among the most efficient deep learning architectures. Rather than simply scaling one dimension up and down as is conventionally done, EfficientNet uses a systematic approach to simultaneously scale these realistic dimensions of neural network size to get the best accuracy with minimum computational cost.

2.3.4 Prov-GigaPath

Prov-GigaPath, a deep-learning framework for ultra-large-scale data processing and learning, that utilizes hierarchical pathways to process data at multiple granularities in parallel allowing both granular and high-level feature extraction. With dynamically adaptable paths and attention properties, ProvGigaPath limits computational load at the expense of high model accuracies on various tasks. The design focuses on scalability and can run on large datasets without hitting the performance wall, ideal for modern high-throughput applications such as video content analysis, massive-scale image processing, and online analytics.

2.3.5 ViT

The Vision Transformer (ViT) by (Dosovitskiy et al.,2020) leverages the transformer architecture, initially designed for natural language processing, to solve computer vision tasks by capturing long-range dependencies through self-attention. Instead of using convolutional layers like CNNs, ViT divides an input image into fixed-size patches, embeds them, and processes these patch embeddings in the transformer encoder, similar to token sequences in NLP. Each image patch is linearly projected, and positional encoding is added to maintain spatial information. This self-attention mechanism enables ViT to compute relationships between all patches, which allows for a comprehensive understanding of the global context of an image rather than focusing only on local areas.

2.3.6 Hybrid Models

Hybrid models in machine learning and deep learning refer to models combining two or more different architectures or approaches to leverage their strengths while mitigating their weaknesses. Hybrid models aim to improve performance, generalizability, or efficiency using complementary techniques. In the context of deep

learning for computer vision, hybrid models often refer to those that combine CNNs with other architectures, such as transformers. Due to the inefficiency of analyzing vast amounts of medical data, traditional methods are time-consuming and, thereby, less accurate in diagnosis. The hybrid approach will increase the accuracy of detecting IDC or non-IDC cases and aid radiologists in breast cancer detection by improving model performance beyond traditional algorithms.

2.4 Model development

We conducted experiments using a CNN model and various hybrid architectures, including EfficientNetB0+ViT, Xception + ViT, and pretrained Prov-GigaPath, to enhance breast cancer detection accuracy.

The CNN model consists of four convolutional blocks with filters `[32, 64, 128, 128]`, each followed by batch normalization, max pooling, and dropout (0.2) for feature extraction and regularization. A fully connected layer with 128 units and dropout (0.5) further processes the extracted features, while a softmax output layer performs binary classification. This model, comprising 1.38 million parameters, effectively captures local patterns in breast cancer images.

The EfficientNetB0+ViT hybrid architecture integrates EfficientNetB0 for feature extraction and ViT for capturing global context. EfficientNetB0 extracts features, which are pooled and passed through transformer encoder layers comprising multi-head attention, feed-forward networks, and residual connections. Dense and dropout layers finalize the output, with the model containing approximately 71 million parameters.

The Xception+ViT hybrid architecture employs Xception's depthwise separable convolutions for efficient feature extraction, followed by ViT layers to capture global dependencies. The architecture includes patch embeddings, positional encodings, and transformer blocks, culminating in a softmax output layer for classification. This model comprises approximately 9.6 million parameters, balancing efficiency and performance.

Prov-GigaPath classifier, a neural network model from hugging face transformer is developed using PyTorch, featuring a multi-layer fully connected architecture tailored for binary classification. The network consists of sequential layers, starting with a 1536-dimensional input, followed by three dense layers with 256, 128, and 64 units, respectively. Each layer is paired with batch normalization, ReLU activation, and dropout for regularization, concluding with a final softmax layer for binary output. The AdamW optimizer is employed with weight decay for enhanced regularization and improved convergence. The training process incorporates early stopping based on validation loss to prevent overfitting, making the model robust for pathology image classification tasks.

2.5 Evaluation Metrics

The common metrics used to evaluate the model performance such as accuracy, precision, recall, F1-score and Confusion Matrix.

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

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

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

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

3. Results

The results are presented using tables and plots. The tables show the complete experimental results on both the CBIS-DDSM and IDC datasets, before and after data augmentation. Plots are provided for the best-performing models: CNN and Xception+ViT on the CBIS-DDSM dataset, and Prov-GigaPath on the IDC dataset.

3.1 Summary of results

The tables below provide insights from the experiments, showing that CNN and Xception+ViT achieved the best performance on the CBIS-DDSM dataset, while Prov-GigaPath excelled on the IDC dataset. Data augmentation slightly improved accuracy for all models and significantly boosted recall, especially for the Xception+ViT hybrid model.

Table 1: All results before Augmentation

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CBIS-DDSM	**CNN**	**96.41**	**98**	**94**	**96**
	Xception +ViT	**96.82**	**98**	**94**	**96**
IDC	CNN	90.51	84	82	83
	Xception + ViT	88.83	81	78	80
	ProvGiga Path	**93.23**	**90.42**	**85.15**	**87.71**
	Efficient NetB4 +ViT	89.78	87	74	80
	Efficient NetB0 +ViT	89.34	83	79	81

Table 2 : Result after Augmentation

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CBIS-DDSM	**CNN**	**97.08**	**98**	**95**	**96**
	Xception+ ViT	**96.8**	**96**	**97**	**96**
	EfficientNet B0+ViT	**97.14**	**96.57**	**96.36**	**96.47**
IDC	Xception+ ViT	88.46	81	77	79
	Efficient-NetB0	**90.67**	**84**	**83**	**83**

3.2 Best performing models

The plots below showcase the best-performing models. CNN demonstrated high performance and generalizability, while Xception+ViT had the lowest misclassification rates based on the confusion matrix. ProvGigaPath outperformed all other models on the IDC dataset across all evaluation metrics.

3.2.1 CNN (CBIS-DDSM)

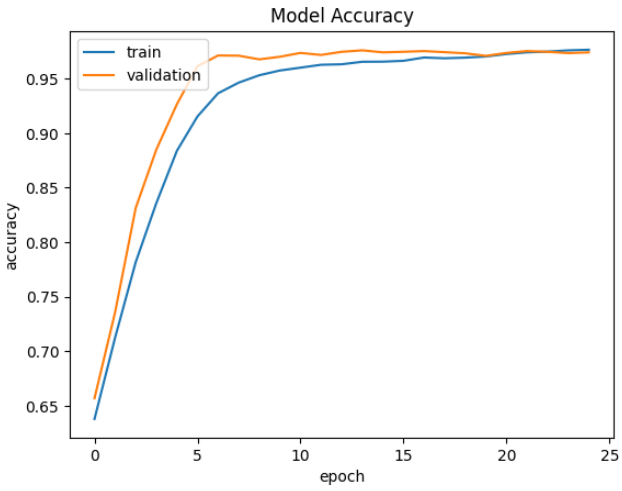

Figure 2: Accuracy curve (CNN)

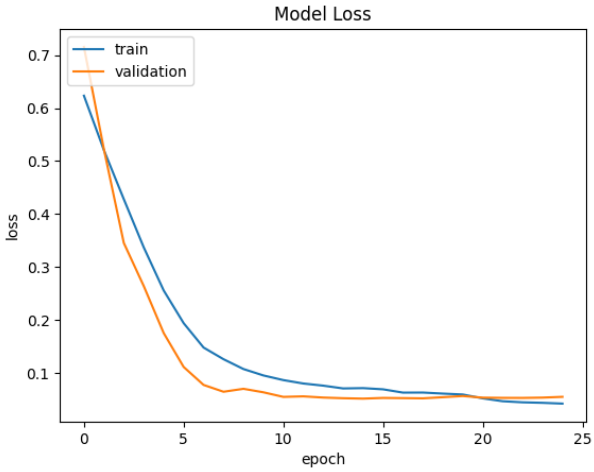

Figure 3: The loss curve (CNN)

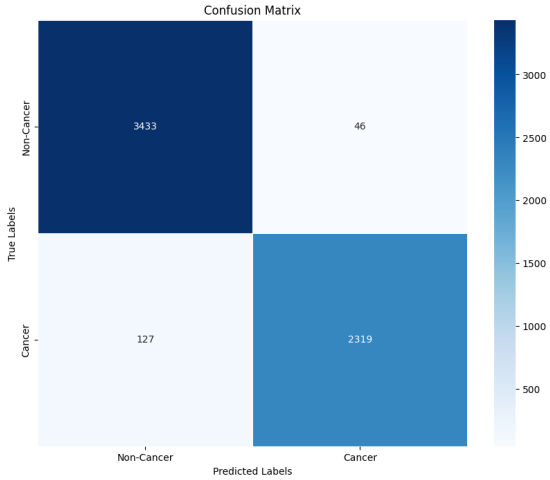

Figure 4: Confusion Matrix (CNN)

3.2.2 Xception+ViT

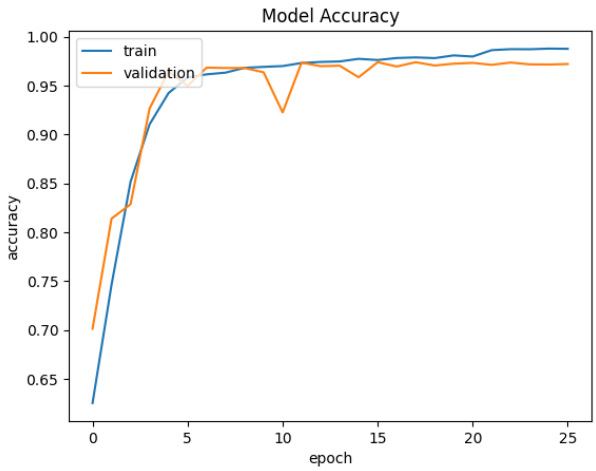

Figure 5: Accuracy curve (Xception+ViT)

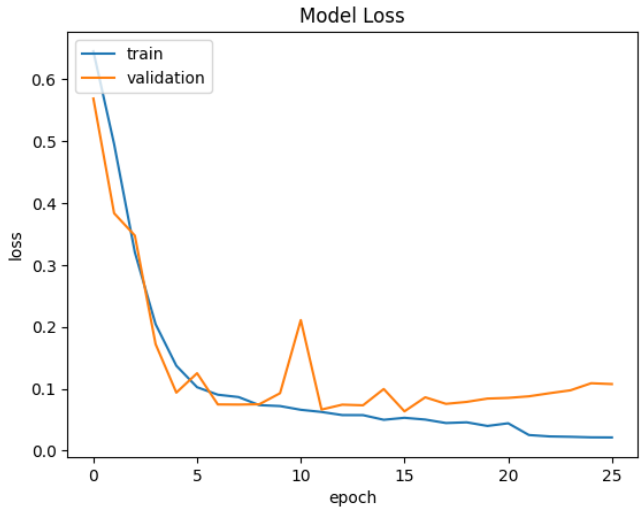

Figure 6 : The Loss curve (Xception+ViT)

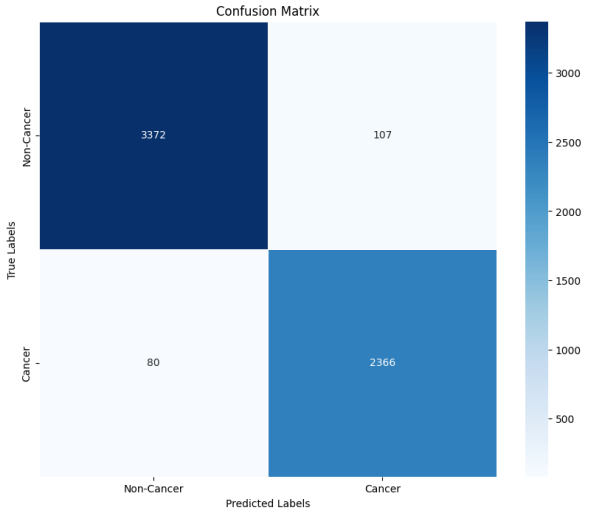

Figure 7 : Confusion Matrix (Xception+ViT)

3.2.3 **ProvGigaPath (IDC)**

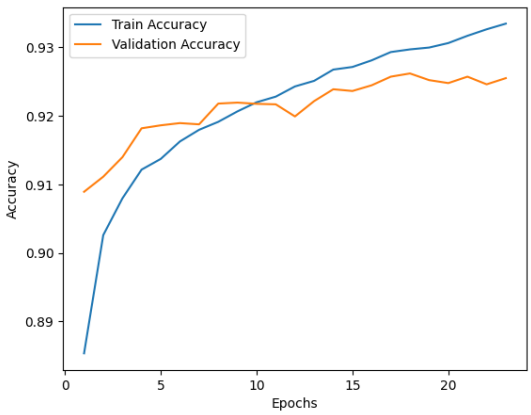

Figure 8 : Accuracy curve (IDC)

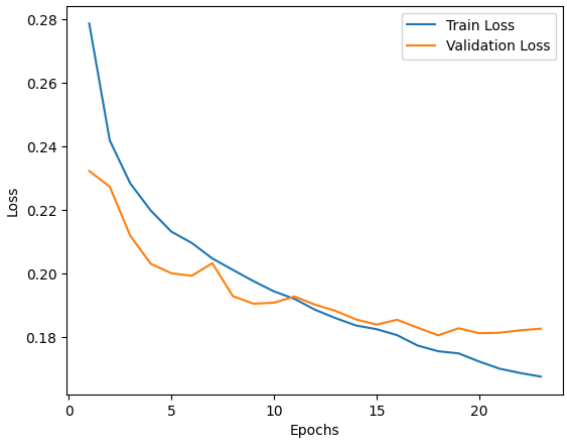

Figure 9 : The Loss curve (IDC)

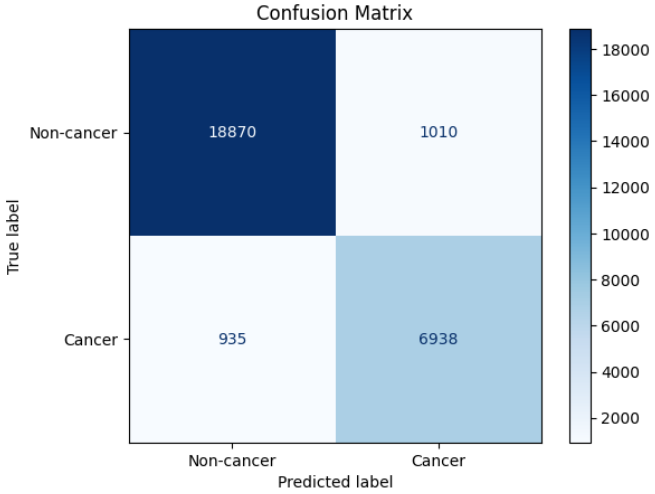

Figure 10 : Confusion matrix (IDC)

4. Discussion

This study highlights the effectiveness of hybrid models, particularly CNNs combined with ViT, for breast cancer detection. The **Xception+ViT** model achieved a **97%** recall on the CBIS-DDSM dataset, which is crucial for minimizing false negatives in cancer detection. Data augmentation techniques like rotation flip and zoom contributed to this success. Similarly, the ProvGigaPath model excelled on the IDC dataset, showcasing strong performance in identifying cancerous cases with minimal misclassification.

However, EfficientNetB0+ViT, using augmentations such as CLAHE, elastic deformation, contrast stretching, and cutout, achieved a validation accuracy of 97.14% and a training accuracy of 88%, which is abnormal. While augmentation was applied only to the training set, improving training accuracy remains a challenge due to the high computational costs of this approach.

5. Conclusion

Xception+ViT achieved the best recall on the CBIS-DDSM dataset, demonstrating the potential of hybrid models in reducing false negatives. However, ProvGigaPath surpassed hybrid approaches on the IDC dataset, achieving 93.23% accuracy. While data augmentation offered slight improvements in generalization, challenges remain in significantly enhancing model performance. Limitations of this study include difficulty in identifying optimal hyperparameters for data augmentation, improvement Prov-GigaPath accuracy and the computational expense of implementing these methods. Future work should focus on integrating CBIS-DDSM and IDC datasets to enhance data diversity and generalization, further fine-tuning to address current gaps, exploring robust preprocessing methods specific to mammography and histopathology, and optimizing augmentation strategies for better results.

6. References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249, 2021.

[2] Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters, 125, 1-6, 2019.

[3] Bolhasani, H., Amjadi, E., Tabatabaeian, M., & Jassbi, S. J. A histopathological image dataset for grading breast invasive ductal carcinomas. Informatics in Medicine Unlocked, 19, 100341, 2020.

[4] Sahu, A., Das, P. K., & Meher, S. (2023). Recent advancements in machine learning and deep learningbased breast cancer detection using mammograms. *Physica Medica*, *114*, 103138.

[5] Wang, L. (2024). Mammography with deep learning for breast cancer detection. Frontiers in Oncology, 14. https://doi.org/10.3389/fonc.2024.1281922

[6] *Breast Cancer Statistics | How Common Is Breast Cancer?* (n.d.). Retrieved September 8, 2024, from https://www.cancer.org/cancer/types/breastcancer/about/how-common-is-breast-cancer.html

[7] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6123282/

[8] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[9] Zuo, D., Yang, L., Jin, Y., Qi, H., Liu, Y., & Ren, L. (2023). Machine learning-based models for the prediction of breast cancer recurrence risk. BMC Medical Informatics and Decision Making, 23(1). https://doi.org/10.1186/s12911-023-02377-z

[10] Mahmood, T., Saba, T., Rehman, A., & Alamri, F. S. (2024). Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. Expert Systems with Applications, 249, 123747.

[11] Park, E. K., Lee, H., Kim, M., Kim, T., Kim, J., Kim, K. H., Kooi, T., Chang, Y., & Ryu, S. (2024). Artificial Intelligence-Powered Imaging Biomarker Based on Mammography for Breast Cancer Risk Prediction. Diagnostics, 14(12), 1212. https://doi.org/10.3390/diagnostics14121212

[12] World Health Organization. (2023, February 3). WHO launches new roadmap on breast cancer. *WHO*. https://www.who.int/news/item/03-02-2023-wholaunches-new-roadmap-on-breast-cancer

[13] Boesch, G. (2024, May 16). *Xception Model: Analyzing Depthwise Separable Convolutions - viso.ai*. Viso.ai. https://viso.ai/deep-learning/xception-model/

[14] Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., ... & Poon, H. (2024). A whole-slide foundation model for digital pathology from real-world data. Nature, 1-8.