Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images

TANIMOLA, Oluwatosin, SHOBAYO, Olamilekan <http://orcid.org/0000-0001-5889-7082>, POPOOLA, Olusogo <http://orcid.org/0000-0002-6026-9816> and OKOYEIGBO, Obinna

analytics

MDPI

Article

Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images

Oluwatosin Tanimola [1], Olamilekan Shobayo [1,*], Olusogo Popoola [1] and Obinna Okoyeigbo [2]

[1] School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield S1 2NU, UK; oluwatosin.s.tanimola@student.shu.ac.uk (O.T.); o.popoola@shu.ac.uk (O.P.)
[2] Department of Engineering, Edge Hill University, Ormskirk L39 4QP, UK; obinna.okoyeigbo@edgehill.ac.uk
* Correspondence: o.shobayo@shu.ac.uk

Abstract: Breast cancer is the most prevalent type of disease among women. It has become one of the foremost causes of death among women globally. Early detection plays a significant role in administering personalized treatment and improving patient outcomes. Mammography procedures are often used to detect early-stage cancer cells. This traditional method of mammography while valuable has limitations in its potential for false positives and negatives, patient discomfort, and radiation exposure. Therefore, there is a probe for more accurate techniques required in detecting breast cancer, leading to exploring the potential of machine learning in the classification of diagnostic images due to its efficiency and accuracy. This study conducted a comparative analysis of pre-trained CNNs (ResNet50 and VGG16) and vision transformers (ViT-base and SWIN transformer) with the inclusion of ViT-base trained from scratch model architectures to effectively classify mammographic breast cancer images into benign and malignant cases. The SWIN transformer exhibits superior performance with 99.9% accuracy and a precision of 99.8%. These findings demonstrate the efficiency of deep learning to accurately classify mammographic breast cancer images for the diagnosis of breast cancer, leading to improvements in patient outcomes.

Keywords: breast cancer; deep learning; transformer models

Citation: Tanimola, O.; Shobayo, O.; Popoola, O.; Okoyeigbo, O. Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images. *Analytics* **2024**, *3*, 461–475. https://doi.org/10.3390/analytics3040026

Academic Editor: Qingshan Jiang

Received: 11 September 2024
Revised: 29 October 2024
Accepted: 7 November 2024
Published: 11 November 2024

1. Introduction

Breast cancer is the most prevalent type of disease among women [1]. It has become the foremost cause of death [2]. Research shows that about 684,996 breast cancer death incidents were recorded in the year 2020 with about 2.3 million newly diagnosed cases [3]. There has been a 0.5% increase in incident rates annually in recent years [4]. The diagnosis of breast cancer is based on the classification of tumours. Breast cancer's early detection and diagnosis are vital for personalized treatment. Early detection can help health professionals and patients discover new treatment options and ensure higher survival rates with a better quality of life [5]. There are two major types of cancer tumours namely, benign and malignant tumours [6]. Early detection of breast cancer is anchored via mammography (MG) [7]. Mammography can detect early-stage breast cancer even when the lump has yet to fully form to a stage where it can be felt as a lump [8]. However, there are occurrences of false diagnoses due to the complexities of MGs [9] which require specialized health professionals to administer them. Due to the difficulty in distinguishing the tumours even by experts, there is a need to automate the diagnostic system [10]. Automated procedures have emerged and have been explored to address the limitation of misdiagnosis. Examples of automated diagnostic procedures include image analysis using computer-aided designs (CADs), biomarker analysis, and Electronic Health Records. Automated procedures aim to assist doctors and health professionals' diagnostic analysis and pattern recognition decisions [11]. However, it poses other challenges like data quality, variability, and issues with human breast complexities. More recently, the deep learning approach has shown a

promising solution as it has shown capability in uncovering intricate features from massive amounts of mammographic images. Deep learning can analyse images at the pixel level, thus enhancing the possibility of detecting important features that other methods may not detect. This ability to detect and extract meaningful information from images makes it a more viable procedure for improving the accuracy of breast cancer diagnosis.

While the use of mammography has been effective in reducing the mortality rate of breast cancers [12], there are major drawbacks to this procedure. Interpreting a mammogram accurately is usually a challenge for many radiologists [13]. Also, analysing tons of mammogram images is not practical for radiologists; the task consumes a lot of time and is exhausting, which leads to false positives or false negatives [14]. According to [13], errors in diagnosis constituted the most common basis of malpractice suits against radiologists. The bulk of such cases were caused by false diagnoses of breast cancer using mammograms [13]. Therefore, there is a probe for more accurate techniques used in detecting breast cancer leading to the use of machine learning in the classification of diagnostic images [14]. Machine learning (ML) algorithms offer promising avenues for enhancing breast cancer detection accuracy and efficiency. Recently deep learning has become a leading tool in various research domains including medical images [15]. A growing body of research supports the potential of ML in breast cancer detection. Studies have shown that ML algorithms can achieve high accuracy in classifying mammograms and other medical images [16,17]. In computer vision, image classification is a critical task with several applications in fields such as autonomous vehicles, surveillance systems, and medical imaging [18]. Furthermore, research suggests that ML can be used to analyse additional data points beyond images, such as patient demographics and genetic information, potentially leading to more comprehensive risk assessment [19]. This study aims to explore the use of computer vision through machine learning algorithms to improve upon the limitations of the traditional analysis of mammograms by the following:

Increasing detection accuracy: Deep learning can analyse large datasets of medical images, potentially identifying subtle patterns missed by human radiologists, leading to higher diagnostic accuracy [20].

Reducing false positives: ML models can be trained to differentiate between benign and malignant lesions, potentially reducing the number of unnecessary biopsies [21]

2. Related Works

The patient's survival rate can be enhanced through early detection and accurate classification of breast cancer [22]. In recent times, deep learning has proven to be a promising method to enhance medical imaging analysis and classification [23]. Ref. [24] conducted a study on the classification of breast cancer histology images using CNNs showing the capacity of CNNs to extract important features from medical image data. The study also used the extracted features to train the Support Vector Machine (SVM) classifier, achieving a comparable result with an 83.3% accuracy and a 95.6% sensitivity to cancer cases [24].

Ref. [25] conducted a review of the classification of mammogram images using the convolution neural network classifier with the K-means clustering techniques compared with conventional networks such as the ANN, SVM, LSVM, and DNN. In his study, the CNN was used to obtain intricate features from mammographic images, a wiener filter was introduced to expel noise, and the K-means clustering techniques were utilized for segmentation. The results achieved with the CNN were better than others with a training accuracy of 96.947% and a testing accuracy of 97.143%, proving the capability of CNNs to efficiently classify breast cancer images. Another study was carried out by [26] in using deep learning for the classification of breast cancer digital mammography. A transfer learning approach was applied to the Inception v3 pre-trained network achieving an accuracy of 88.2%.

Research conducted by [27] applied deep learning in mammography image segmentation and classification with an automated CNN approach. The study objective is to explore

robust deep learning models that have a good performance in computer vision. The study used Inception V3, DenseNet121, ResNet50, VGG16, and MobileNetV2, and the models were applied to three different mammographic image datasets from MIAS, DDSM, and CBIS-DDSM databases. The study utilized the U-Net-modified segmentation model for the segmentation process. The research achieved the best result on the DDSM database with the InceptionV3 model achieving the best model performance with 98.8% accuracy, 98.88% AUC, 98.79 precision, and an F1 score of 97.99%.

In a recent study, ref. [28] introduced a multi-class classification of breast cancer from the histopathological image datasets utilizing the ensemble of SWIN transformers. They investigated with the ensemble of the SWIN transformer to classify benign and malignant cases, a two-class classification versus an eight-class classification of four benign and four malignant cases of breast cancer on histopathology images. The study achieved a result of 96.0% on the eight-class classification and 99.6% for the two-class classification. The SWIN transformer, a variant of the vision transformer that works on the concept of non-overlapping shifted windows [29], is a proven method for computer vision tasks. The authors in [30] classified breast cancer on thermograms using a hybrid model of deep CNNs and transformers. The study utilized the TransUNet, a variant of the vision transformer for the segmentation process, and four CNN model architectures were utilized, namely EfficientNet-B7, ResNet-50, VGG-16, and DenseNet-201, for classification into three categories of healthy, sick, and unknown. ResNet50 achieved the best performance with 97.26% accuracy and 97.26% sensitivity.

A study on mammography breast cancer classification using vision transformers was conducted by [31] using a pre-trained vision transformer on a mammography breast image for the classification of benign and malignant cases. The model utilized the DDSM dataset and achieved a result of 99.96% accuracy, 99.95% precision, and 99.96% sensitivity for the breast cancer classification. This study focuses on a comparative analysis of chosen CNN-based architecture, specifically the ResNet50 and the VGG16. This is because they have proven to achieve good performance on the medical image datasets being trained on a large image dataset (ImageNet) and have been previously utilized on deep learning medical imaging tasks with good results. This is in comparison to the novel vision transformer introduced by [32]. Vision transformers have become the preferred model in medical imaging because of their exceptional performance. The study utilized the ViT-base and the SWIN-b transformer model. Also, the ViT was trained from scratch on the mammogram dataset to show how pre-trained transformer models perform over models training from scratch on medical images.

3. Materials and Methods

3.1. Data Collection

The dataset used for the study was obtained from the Mendeley dataset archive [33] for breast mammography images with masses comprising a total of 24,576 images collected from the INbreast dataset with 7632 images, the MIAS dataset with 3816 images, and the DDSM dataset with 13,128 images. These images were aggregated together to give the 24,576 images used for training the models. The images were all in size 227 × 227 pixels. To prepare the images for training purposes, they were preprocessed, resized, augmented, and normalized. The dataset type and its distribution are presented in Table 1 below.

Table 1. Summary of breast images from MIAS, DDSM, and IN Breast datasets available for training and testing.

Dataset Type	No. of Benign	No. of Malignant	Total Images
MIAS	2376	1440	3816
DDSM	5970	7158	13,128
IN Breast	2520	5112	7632
	TOTAL IMAGES		24,576

3.2. Shifted Window (SWIN) Transformer

We propose the use of a SWIN transformer for the classification of breast cancer in this study. The introduction of the SWIN transformer by [29] attempts to solve the limitations posed by the vision transformer. This limitation is related to how the vision transformer extracts patches using the 16 × 16 patch size to keep the image sequence as introduced by [32]. This is more efficient with low- to medium-sized pixel images; however, many computing vision tasks require detailed information at the pixel level, such as semantic segmentation, which is also useful in medical imaging requiring a heavy prediction at the image pixel level. Furthermore, high-resolution images will result in computation inefficiencies [29]. The semantic SWIN transformer utilizes a 4 × 4 patch size at the initial division where each patch is treated as a token on each dimension of the RGB image giving a total of 4 × 4 × 3 which equals 48, as illustrated by the formula for stage 1.

$$\frac{H}{4} \times \frac{W}{4} = 48 \tag{1}$$

Then, a linear transformation will convert each patch into a C-dimension vector, and this is processed by the transformer blocks. The SWIN transformer block consists of a shifted window multi-head self-attention module as a replacement for the vision transformer standard multi-head self-attention module [29].

As given in Figure 1, the hierarchical representation construction begins from small patches and then systemically merges neighbouring patches in subsequent deeper transformer layers [29]. The computation of self-attention included a relative position bias procedure to each head in computing similarity, following a similar work from [29,34,35].

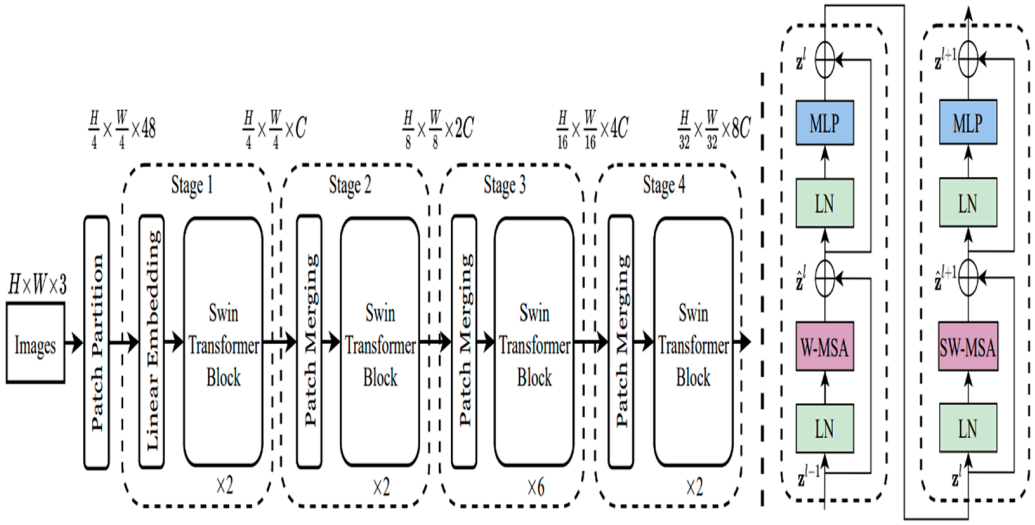

Figure 1. SWIN transformer architecture with equation [29].

The mathematical equation for the attention layer including a relative position bias B $\in R^{M\,2 \times M\,2}$ is as follows:

$$\text{Attention } ((Q, K, V) = \text{SoftMax } (QK^T / \sqrt{d} + B)\, V \tag{2}$$

where Q, K, and V $\in R^{M\,2 \times d}$ are the query, key, and value matrices; d is the query/key dimension, and M^2 is the number of patches in a window [29].

3.3. Model Development

The developmental framework for the experiment is shown in Figure 2.

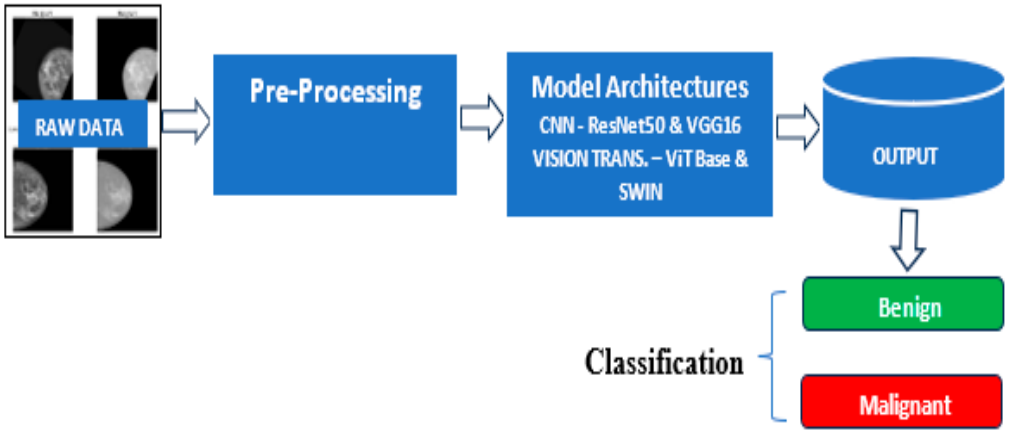

Figure 2. Experimental development framework.

A comparative analysis of the two major frameworks for image classification was carried out in this study. Convolution neural networks (CNNs) and vision transformers were used as a framework for training the models. Specifically, VGG16 [36] and ResNet50 [37], both pre-trained, were applied for the CNN models and ViT-base, and SWIN transformer architectures [32] were utilized for the vision transformer models. Based on the insight derived from a review of the literature, these models have proven to perform excellently in image classification and are suitable for medical image analysis [37]. These models were fine-tuned on pre-trained architectures with the vast number of parameters making it suitable for learning various patterns in image classification. Python libraries such as TensorFlow, Pytorch, Keras, and OpenCV were leveraged to develop these architectures. Adam and SGD optimizers were used to optimize the model parameters during training. Hyperparameters such as learning rate, batch sizes, and the number of epochs were optimized using grid search and random search.

3.3.1. Data Augmentation and Preprocessing Considerations

Data Preprocessing

Normalization: To facilitate a consistent range of pixel values, a normalization technique was adopted to scale the pixel values between 0 and 1. This is aimed at optimizing the model performance and improving the model convergence. The images were normalized for the study using a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. This aligns the pixel values with the pre-trained SWIN transformer model distribution.

Resizing: The mammographic images were resized into a standard dimension of 227 by 227 pixels. This makes the dataset suitable for analysis and training by ensuring consistency of the image sizing.

Data Augmentation Technique

Rotation: Random rotations were applied to images to introduce variability in a specific range. This will help the model to generalize well and improve efficiency.

Shifting: Additional shifts by *n* values along the x and y axes were applied to augment the data when the mass was in different positions within the image, which helped the model to detect masses in different locations.

Brightness Adjustment: Another technique used in the approach was to use images of the same or similar objects in different lighting conditions with different levels of brightness.

These chosen augmentation techniques aimed to mimic real environment variation and improve the model's robustness to variability in imaging conditions.

3.3.2. Class Distribution

After the preprocessing and augmentation process, the dataset is analysed for class imbalance. The distribution of classes was performed on both the training and test data after a train–test split ratio of 80:20. The distribution of the classes is shown in Figures 3 and 4.

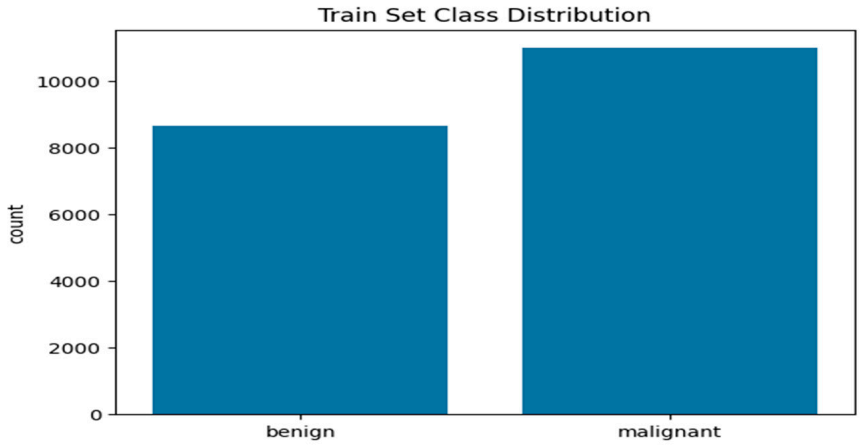

Figure 3. Plot of class distribution of training dataset.

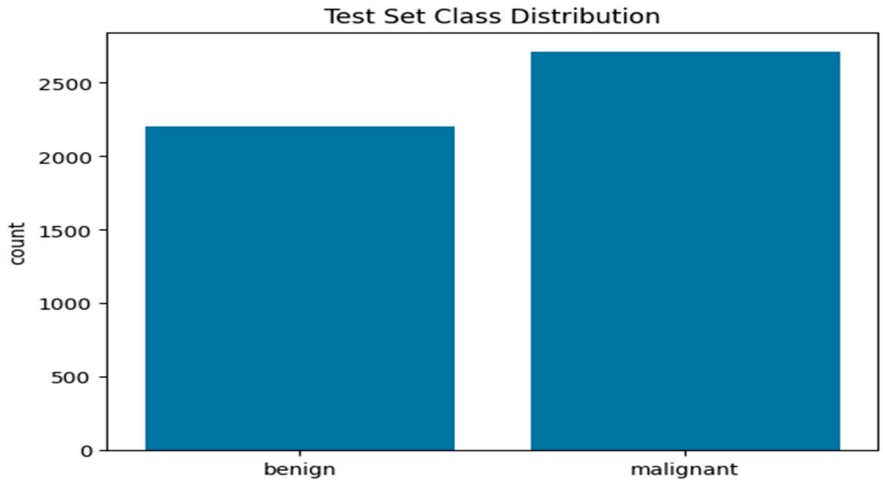

Figure 4. Plot of class distribution of training dataset.

Figure 3 shows the histogram plot of the training set class distribution. The plot shows 11,000 malignant cases and 8660 benign cases that make up the training set, indicating a ratio of 1:1.27 of malignant and benign cases, respectively. This implies that, for every 100 benign cases, there are approximately 127 malignant cases.

Figure 4 shows the histogram plot of the testing set class distribution. The plot shows 2206 benign cases and 2709 malignant cases for the testing dataset, indicating a ratio of 1:0.813 for malignant and benign cases. Since the difference in class between the two cases is relatively low. It can be considered balanced. Figure 5 shows random samples of the images in the dataset, with augmentation.

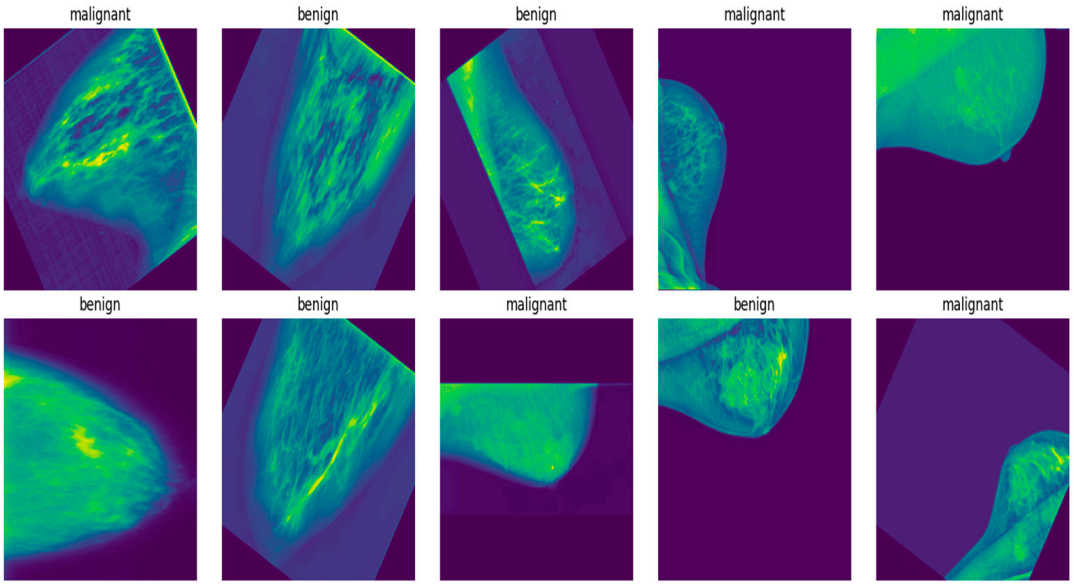

Figure 5. Random samples of mammogram image dataset.

3.4. Experimental Set-Up

The Google Collab platform was chosen to deploy the model, due to its enhanced computational efficiencies. The specification of the computational resources utilized to train our model within the Google Collab runtime is an NVIDIA L4 GPU with a memory capacity of 23,034 MiB. To obtain the dataset loaded onto the Google Collab platform, there is a need for the organization of the dataset. This was highlighted by [38] in their work on melanoma classification on dermoscopy images using a neural network ensemble model by organizing the dataset into directories representing classes of images. This procedure simplifies the data-loading process and makes accessing the directories straightforward. The method used aligns with the data-loading procedure of [38]; this was adopted because of its simple and compatible approach to dataset loading. The dataset organization into directories was carried out using Python. The base directory called total masses consists of two sub-directories called the train and test. Within the train and test sub-directories are the classified masses labelled malignant and benign which contain respective mass images. Some important embedded functions for the loading and preprocessing of images were utilized. These are robust deep learning frameworks such as TensorFlow and PyTorch [39]. The capabilities of these libraries can handle large datasets and improve computational processes [40].

4. Results

4.1. Model Training

A comparative analysis study was carried out using pre-trained models of ResNet50 [41], VGG16 [36], ViT-base [32], and SWIN transformer [29] architectures. These models were originally trained on ImageNet datasets. They were fine-tuned on the breast cancer image dataset consisting of 24,576 images to classify benign and malignant cases.

The parameters we used to fine-tune the CNN and transformer models are given in Table 2.

The model was trained for 10 epochs while simultaneously being validated on the testing set. A plot of the performance was generated to show the training and testing accuracy during the training process. The code for the model development can be accessed via the following GitHub repository: https://github.com/tossign/Breast-cancer-classification/tree/main/models (accessed on 10 September 2024).

Table 2. Model parameters for fine-tuning.

Model	Loss Function	Optimizer	LR	Batch Number	No. of Epochs
SWIN	Cross Entropy Loss	Adam	1×10^{-4}	64	10
ViT-Base	Cross Entropy Loss	Adam	1×10^{-3}	32	10
VGG16	Cross Entropy Loss	Adam	1×10^{-5}	64	10
ResNet50	Cross Entropy Loss	Adam	1×10^{-3}	64	10

Figure 6 shows the training loss and accuracy performance for the ViT-base pre-trained model. As illustrated, the training and test loss steadily declined across the 10 epochs. While the test loss lags behind the training loss initially until the sixth epoch, it interestingly picks up and is slightly higher than the training loss at the sixth and eighth epochs. Similarly, the training and test accuracies increase steadily across the training epochs Figure 6 with a slight difference in the training and testing accuracies until the 6th and 8th epochs when the test accuracy slightly decreases below the training accuracy but eventually converges well at the 10th epoch. This shows that the model effectively learns from the training data and appreciably generalizes. However, further investigation may be conducted to ensure no trace of overfitting and improve the model's generalization performance.

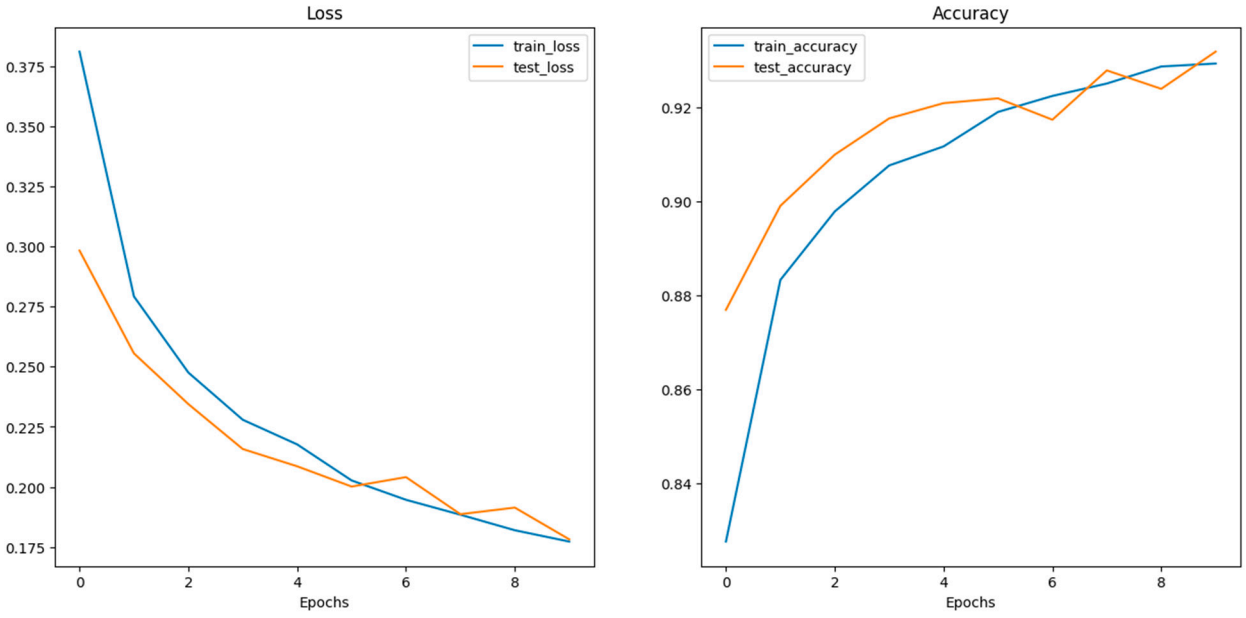

Figure 6. Performance plot of training loss and accuracy for ViT-base pre-trained model.

The training and testing loss of the SWIN transformer as shown in the Figure 7 plot above indicates that the training and testing loss declines steadily throughout the training epochs. There is a relatively very small gap between the training and testing curve, indicating that the model generalizes significantly well with unseen data, learning discrete features of the data. Also, the training and testing accuracy increases steadily throughout the training epochs. Also, the gap between them is relatively small, indicating a good generalization of the model with the data. The training and testing accuracy achieved a convergence point indicating the model's optimal performance. Therefore, the SWIN transformer architecture performs exceptionally on the given data with a very high training and testing accuracy of 0.9981 and 0.994, respectively.

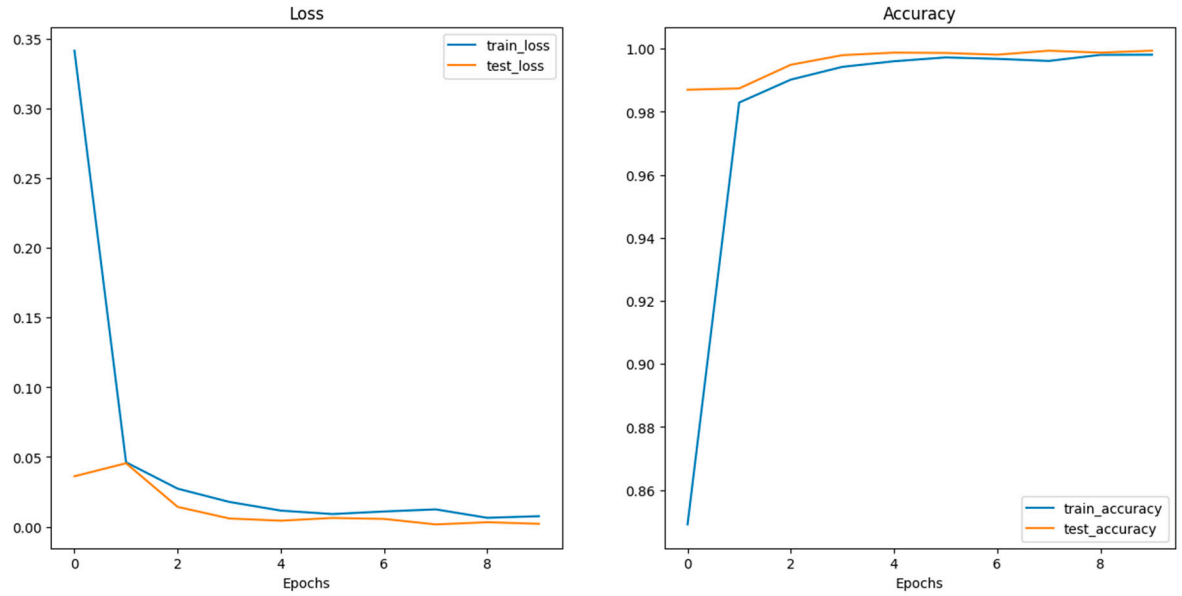

Figure 7. Performance plot of training loss and accuracy for SWIN transformer pre-trained model.

The training and testing loss for the ResNet50 pre-trained model decreases across the 10 epochs as shown in Figure 8. The test loss decreases and lags behind the training loss until it reaches its lowest point. Interestingly, there was a sharp surge at epoch 4 and epoch 7 while the training loss also surged higher after the convergence at epoch 8. The minimum training loss is about 0.3947 and the test loss at 0.3907. Meanwhile, the training accuracies increase steadily across the epochs until a slight drop at the eighth epoch, while the overall accuracy trend is upward, and the test accuracy does not give a monotonic pattern as it fluctuates across the epochs. This may be due to variability in the accuracy, and the model may not be sensitive to variations in the data causing the fluctuations. While this may not pose any significant concerns, further investigation such as hyperparameter tuning, training on more epochs, and regularization techniques may be employed. The test accuracy peaks at 0.8277 while the training accuracy peaks at 0.8194.

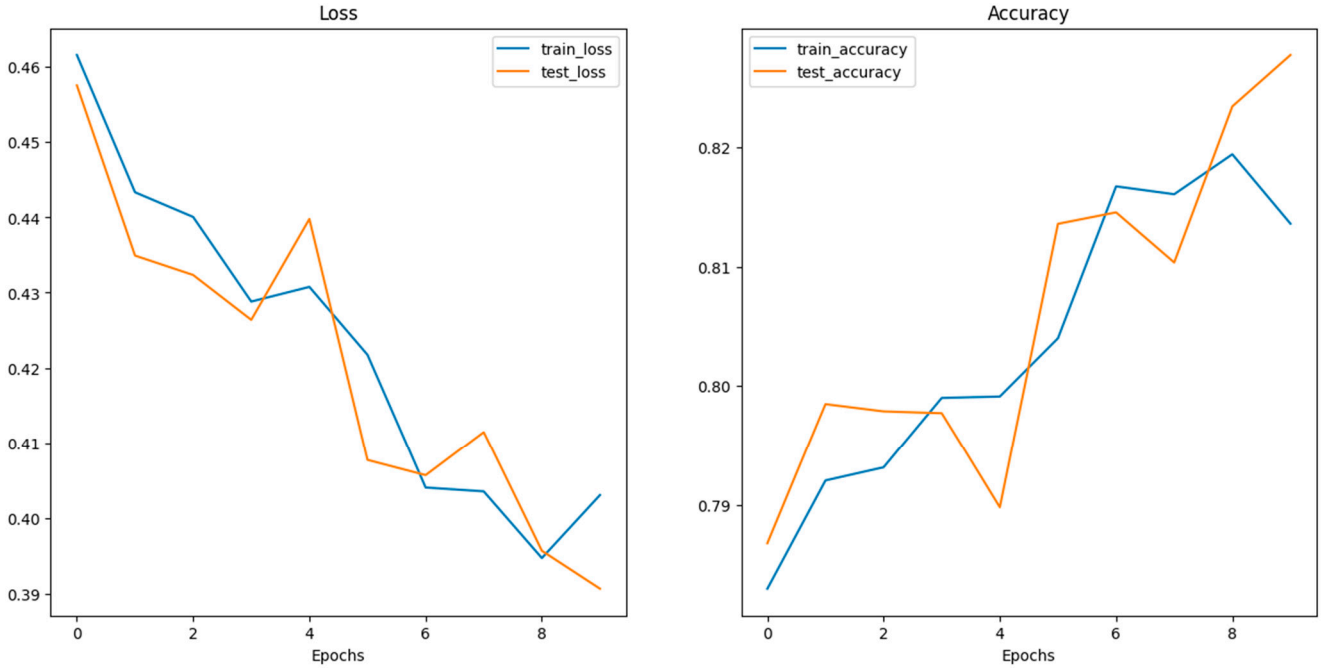

Figure 8. Performance plot of training loss and accuracy for ResNet50 pre-trained model.

The training and testing loss plot as shown in Figure 9 shows a decreasing trend indicating that the model is still learning while that gap between the training and testing loss may be an indication that the model is learning well but struggling to generalize well with unseen data. This may pose a possibility of overfitting in the model. This may be investigated more by allowing the model to train more across longer epochs to understand the model performance better. Similarly, training and testing accuracy continue to increase but at a slower rate in earlier epochs. Training on more epochs may give more gains as the training and test may reach a convergence point. It appears the test loss and the test accuracy reach a near-constant level, indicating that the model may be struggling to generalize well and is less sensitive to data variability. However, further hyperparameter tuning and an increase in epochs may improve model performance.

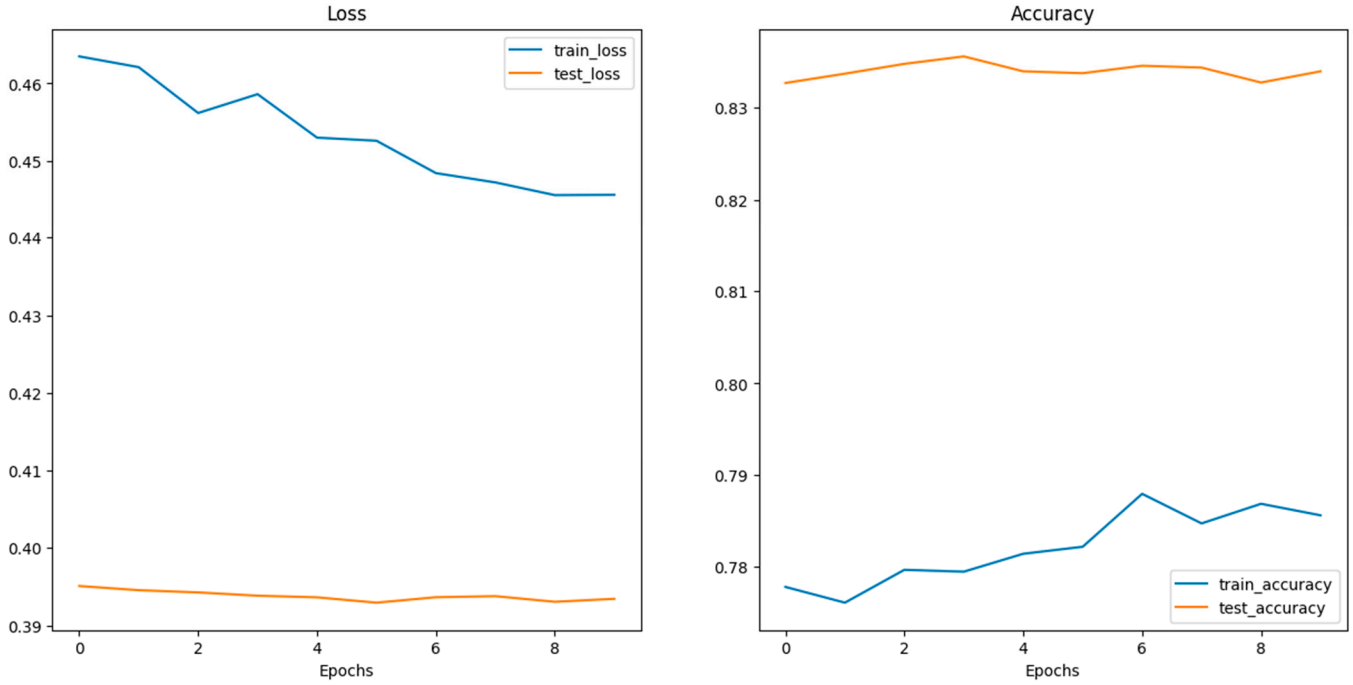

Figure 9. Performance plot of training loss and accuracy for VGG16 pre-trained model.

4.2. Confusion Matrix

Based on the confusion matrices shown in Figures 10–13. The SWIN transformer architecture shows exceptional performance with 2203 True Negatives and 2709 True Positives as shown in Figure 13. This makes it the best-performing model having zero False Positives and only three False Negatives. The ViT-base comes close with 2011 True Negatives and 2568 True Positives and has 141 False Negatives and 195 False Positives as shown in Figure 10.

$$\text{Accuracy} = \frac{Correctly\ classified\ classes}{Total\ classification} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{Correctly\ classified\ True\ positives}{All\ positive\ classified} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Specificity} = \frac{True\ Negatives}{All\ actual\ negatives\ (including\ mis-classified)} = \frac{TN}{TN + FP} \tag{5}$$

$$\text{Sensitivity} = \frac{Accurately\ classified\ actual\ positives}{All\ actual\ positives\ (including\ mis-classified)} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN+FN} \quad (7)$$

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP+FP} \quad (8)$$

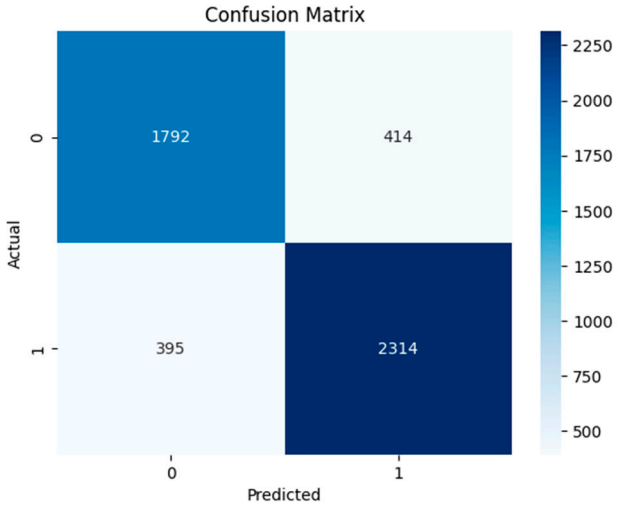

Figure 10. Confusion matrix of VGG16.

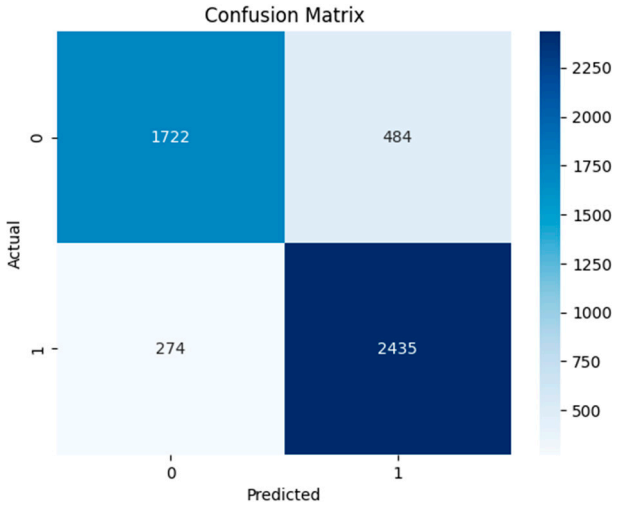

Figure 11. Confusion matrix of ResNet50.

The VGG16 architecture has 1792 True Negatives and 2394 True Positives while having 395 False Negatives and 414 False Positives as shown in Figure 8. The ResNet50 architecture shows 1722 True Negatives and 2435 True Positives while having 274 False Negatives and 484 False Positives as shown in Figure 9.

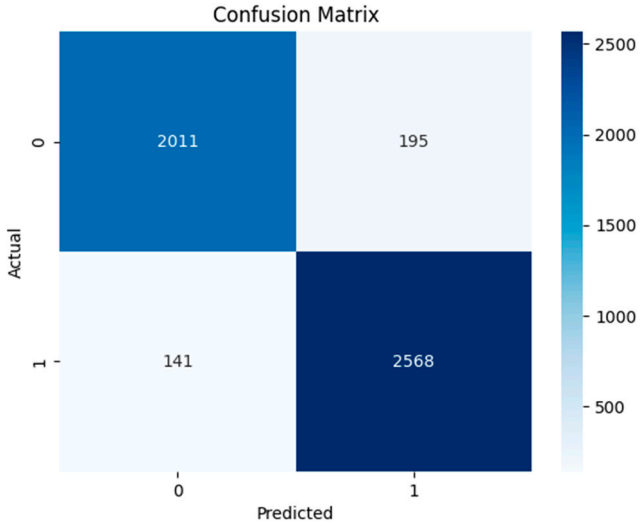

Figure 12. Confusion matrix of ViT-base.

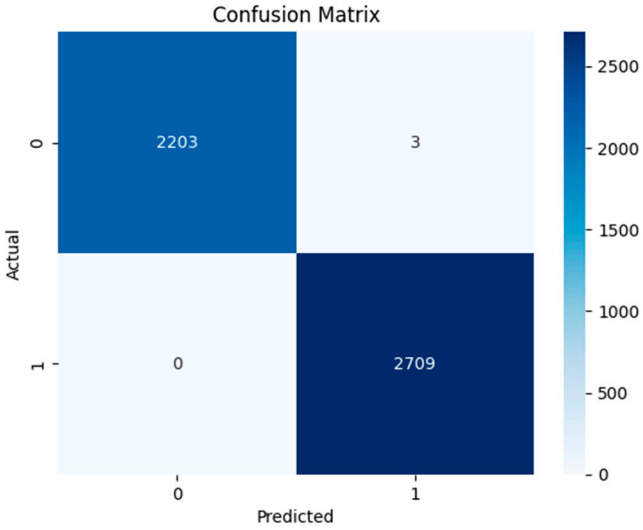

Figure 13. Confusion matrix of SWIN transformer.

Table 3 shows the sensitivity and specificity values of all the model architectures. The sensitivity measures the value of the actual positives that the model predicted correctly, while the specificity refers to the model's ability to predict actual negatives correctly. The table shows ResNet50 to have 0.858 sensitivity and 0.794 specificity, while VGG16 has 0.854 sensitivity and 0.811 specificity. ViT-base has 0.948 sensitivity and 0.913 specificity, the SWIN transformer has 1.000 sensitivity and 0.9986 specificity. The negative predicted value (NPV) is the probability of the model truly predicting a negative case for a negative classification. While all the models except the ViT trained from scratch achieve a relatively high value, the SWIN transformer achieved the highest score and a perfect score of predicting a truly negative class as negative, in the benign case. The positive predicted value (PPV) is the probability of the model truly predicting a positive for a given positive classification. Again, the SWIN transformer excelled among the models with a score of 0.9989 indicating its ability to predict positive classification as positive, in the malignant case.

Table 3. Evaluation metrics of model architectures.

Model	Sensitivity (TPR)	Specificity (TNR)	NPV	PPV	Precision	Accuracy
ResNet50	0.858	0.794	0.863	0.834	0.837	0.829
VGG16	0.854	0.811	0.819	0.853	0.848	0.835
ViT-base	0.948	0.913	0.934	0.929	0.929	0.932
SWIN	1.000	0.999	1.000	0.999	0.998	0.998
ViT scratch	1.000	1.000	0	0.551	0.551	0.551

The SWIN transformer displayed the best performance over all the evaluation metrics with excellent scores. The results imply that the model can identify discriminative features during training and effectively and accurately classify breast cancer as benign and malignant. It achieved a training accuracy of up to 0.9981 and a test accuracy of up to 0.9994. The vision transformer (ViT-base) performs very well with high accuracy, low loss, and a fast convergence rate, making the model exhibit good generalization ability. Though the result across all metrics is not as good as the SWIN transformer, its performance is excellent, achieving the highest training accuracy of 0.9293 and test accuracy of up to 0.9318. The result from the ResNet50 architecture shows promising performance with good accuracy, precision, and recall measures. The model builds on feature extraction through the pre-trained weights from ResNet50. The resulting training accuracy is up to 0.8194 and the test accuracy is up to 0.8277. The result from the VGG16 arch accuracy shows a promising performance slightly better than the ResNet50 in the accuracy score. With good accuracy, precision, and recall measures, VGG16 can classify medical images accurately. The model builds on feature extraction through the pre-trained weights from VGG16, achieving a training accuracy of up to 0.7972 and a test accuracy of up to 0.8264. The result obtained from the scratch training of the vision transformer lags behind the other four architectures exhibiting relatively low accuracy, precision, recall, and F1 scores in comparison with other architectures. The training accuracy is up to 0.5595, while it achieved a test accuracy that is consistently around the value of 0.5524, suggesting a limitation in the model's learnability of generalizable features. This is largely due to a limited dataset. An effective deep learning model requires a large amount of data for training with a huge number of parameters [4], hence the superior performance of the transfer learning on pre-trained models over the trained models from scratch. The model does not appropriately learn and generalize well with such a limited image. This is in comparison with pre-trained models that have been subjected to large and complex image data.

5. Conclusions

In this study, we made a comparison between the CNN-based architectures and Vision transformer-based deep learning models to identify the best-performing model to accurately classify mammographic breast cancer images. The SWIN transformer model achieved the best evaluation results in terms of accuracy and precision. This attests to the potential and effectiveness of transformer architectures in computer vision tasks, which could be influential in the early detection and diagnosis of breast cancer, offering a practical application to improve patient outcomes and offer valuable tools for healthcare professionals. The same could be said about the pre-trained ViT model, offering quite impressive valuation results as well. This can be attributed to the long-range dependency feature retention capability of their architecture. Using pre-trained transformer models will provide better performance due to the vast learnable parameters with weights updated through learning from millions of images, outperforming CNN models which used to be the leader in computer vision tasks. For future research, to gain explainability of the transformer architecture, there is a need to obtain larger mammograms to train the transformer from scratch as results have shown that models give better performances with larger datasets.

Also, other medical imaging technologies such as ultrasound and MRI alongside mammogram images can be used for training the models for a more robust, effective, and improved classification model. This is particularly beneficial as each imaging procedure offers unique perspectives on breast tissues. Combining these unique features can improve the performance and accuracy of the deep learning model by giving it a better generalization of diverse breast tissues. The classifier can be enhanced to provide features to support patients in terms of medical information, referrals to medical practitioners, and prescriptions, as this can provide bespoke treatments for patients who have been diagnosed with the disease.

Author Contributions: Conceptualization, O.T. and O.S.; methodology, O.T. and O.S.; software, O.T. and O.S.; validation, O.S., O.P. and O.O.; formal analysis, O.T. and O.S.; investigation, O.T. and O.S.; resources, O.S.; data curation, O.T.; writing—original draft preparation, O.T. and O.S.; writing—review and editing, O.S., O.P. and O.O.; supervision, O.S.; project administration, O.P. and O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The dataset is available at https://doi.org/10.17632/ywsbh3ndr8.2 (accessed on 17 July 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jabeen, K.; Khan, M.A.; Balili, J.; Alhaisoni, M.; Almujally, N.A.; Alrashidi, H.; Tariq, U.; Cha, J. BC(2)NetRF: Breast Cancer Classification from Mammogram Images Using Enhanced Deep Learning Features and Equilibrium-Jaya Controlled Regula Falsi-Based Features Selection. *Diagnostics* **2023**, *13*, 1238. [CrossRef] [PubMed]
2. Nasser, M.; Yusof, U.K. Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction. *Diagnostics* **2023**, *13*, 161. [CrossRef] [PubMed]
3. 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 Cancer. J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
4. Husen, N.; Habtamu, B.; Barki, H.; Choe, S.; Mulugeta, F.; Amdissa, D.; Ayana, G.; Dese, K.; Dereje, Y.; Kebede, Y. Vision-Transformer-Based Transfer Learning for Mammogram Classification. *Diagnostics* **2023**, *13*, 178. [CrossRef]
5. Mohammed, M.; Ikotun, A.M.; Mohamed, T.I.A.; Ezugwu, A.E.; Fonou-Dombeu, J.V. A bio-inspired convolution neural network architecture for automatic breast cancer detection and classification using RNA-Seq gene expression data. *Sci. Rep.* **2023**, *13*, 14644. [CrossRef]
6. Gayathri, B.M.; Sumathi, C.P.; Santhanam, T. Breast Cancer Diagnosis Using Machine Learning Algorithms—A Survey. *Int. J. Distrib. Parallel Syst.* **2013**, *4*, 105–112. [CrossRef]
7. Weedon-Fekjær, H.; Romundstad, P.R.; Vatten, L.J. Modern mammography screening and breast cancer mortality: Population study. *BMJ* **2014**, *348*, g3701. [CrossRef]
8. Pashayan, N.; Antoniou, A.C.; Ivanus, U.; Esserman, L.J.; Easton, D.F.; French, D.; Sroczynski, G.; Hall, P.; Cuzick, J.; Evans, D.G. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat. Rev. Clin. Oncol.* **2020**, *17*, 687–705. [CrossRef]
9. Chougrad, H.; Zouaki, H.; Alheyane, O. Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing* **2020**, *392*, 168–180. [CrossRef]
10. Zhou, Z. Breast Cancer Diagnosis with Machine Learning. *Highlights Sci. Eng. Technol.* **2022**, *9*, 73–75. [CrossRef]
11. Übeyli, E.D. Implementing automated diagnostic systems for breast cancer detection. *Expert Syst. Appl.* **2007**, *33*, 1054–1062. [CrossRef]
12. Heywang-Köbrunner, S.H.; Hacker, A.; Sedlacek, S. Advantages and disadvantages of mammography screening. *Breast Care* **2011**, *6*, 199–207. [CrossRef] [PubMed]
13. Whang, J.S.; Baker, S.R.; Patel, R.; Luk, L.; Castro, A., III. The causes of medical malpractice suits against radiologists in the United States. *Radiology* **2013**, *266*, 548–554. [CrossRef] [PubMed]
14. Zebari, D.A.; Zeebaree, D.Q.; Abdulazeez, A.M.; Haron, H.; Abdul Hamed, H.N. Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access* **2020**, *8*, 203097–203116. [CrossRef]
15. Gheflati, B.; Rivaz, H. Vision Transformers for Classification of Breast Ultrasound Images. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022.

16. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]
17. Liang, T.; Shen, J.; Wang, J.; Liao, W.; Zhang, Z.; Liu, J.; Feng, Z.; Pei, S.; Liu, K. Ultrasound-based prediction of preoperative core biopsy categories in solid breast tumor using machine learning. *Quant. Imaging Med. Surg.* **2023**, *13*, 2634–2646. [CrossRef]
18. Baroni, G.L.; Rasotto, L.; Roitero, K.; Tulisso, A.; Di Loreto, C.; Della Mea, V. Optimizing Vision Transformers for Histopathology: Pretraining and Normalization in Breast Cancer Classification. *J. Imaging* **2024**, *10*, 108. [CrossRef]
19. Alakwaa, W.; Nassef, M.; Badr, A. Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN). *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 409–417. [CrossRef]
20. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef]
21. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. Available online: https://www.sciencedirect.com/science/article/pii/S1361841517301135 (accessed on 10 September 2024). [CrossRef]
22. Cheng, H.; Shan, J.; Ju, W.; Guo, Y.; Zhang, L. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognit.* **2010**, *43*, 299–317. [CrossRef]
23. Mahoro, E.; Akhloufi, M.A. Applying Deep Learning for Breast Cancer Detection in Radiology. *Curr. Oncol.* **2022**, *29*, 8767–8793. [CrossRef] [PubMed]
24. Araújo, T.; Aresta, G.; Castro, E.; Rouco, J.; Aguiar, P.; Eloy, C.; Polónia, A.; Campilho, A. Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* **2017**, *12*, e0177544. [CrossRef] [PubMed]
25. Albalawi, U.; Manimurugan, S.; Varatharajan, R. Classification of breast cancer mammogram images using convolution neural network. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e5803. [CrossRef]
26. López-Cabrera, J.D.; Rodríguez, L.A.L.; Pérez-Díaz, M. Classification of breast cancer from digital mammography using deep learning. *Intel. Artif.* **2020**, *23*, 56–66. [CrossRef]
27. Salama, W.M.; Aly, M.H. Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alex. Eng. J.* **2021**, *60*, 4701–4709. [CrossRef]
28. Tummala, S.; Kim, J.; Kadry, S. BreaST-Net: Multi-class classification of breast cancer from histopathological images using ensemble of SWIN transformers. *Mathematics* **2022**, *10*, 4109. [CrossRef]
29. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. In SWIN Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
30. Mahoro, E.; Akhloufi, M.A. Breast cancer classification on thermograms using deep CNN and transformers. *Quant. InfraRed Thermogr. J.* **2024**, *21*, 30–49. [CrossRef]
31. Abimouloud, M.L.; Bensid, K.; Elleuch, M.; Aiadi, O.; Kherallah, M. Mammography breast cancer classification using vision transformers. In Proceedings of the International Conference on Intelligent Systems Design and Applications, Olten, Switzerland, 11–13 December 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 452–461.
32. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
33. Lin, T.; Huang, M. Dataset of Breast mammography images with Masses. *Data Brief* **2020**, *31*, 105928. [CrossRef]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
35. Bao, H.; Dong, L.; Wei, F.; Wang, W.; Yang, N.; Liu, X.; Wang, Y.; Gao, J.; Piao, S.; Zhou, M. Unilmv2: Pseudo-masked language models for unified language model pre-training. In Proceedings of the International Conference on Machine Learning, PMLR, Online, 13–18 July 2020; pp. 642–652.
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
38. Xie, F.; Fan, H.; Li, Y.; Jiang, Z.; Meng, R.; Bovik, A. Melanoma classification on dermoscopy images using a neural network ensemble model. *IEEE Trans. Med. Imaging* **2016**, *36*, 849–858. [CrossRef] [PubMed]
39. Chollet, F. *Deep Learning with Python*; Simon and Schuster: London, UK, 2021.
40. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.P.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **1912**, *32*, 8026.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *CoRR* **2015**, abs/1512.03385. Available online: http://arxiv.org/abs/1512.03385 (accessed on 10 September 2024).