Breast Cancer Multi-Class Classification using ViT Model

Esam Mohammed Othman
Computer Science Department
College of Applied Computer Science, King Saud University
Riyadh, KSA

ABSTRACT
Breast cancer ranks as the most prevalent form of cancer among women worldwide, underscoring the importance of early detection for enhancing treatment success rates. The ability to accurately differentiate between malignant (aggressive) and benign breast tumors is crucial for determining appropriate treatment strategies. This research introduces a novel methodology leveraging Transformer models for the task of breast cancer image classification. Utilizing a Vision Transformer (ViT) pre-trained across a broad array of domains, this approach incorporates an ensemble of densely connected network layers specifically refined for a dataset dedicated to breast cancer imagery. The performance of this innovative model was rigorously evaluated against a benchmark dataset, demonstrating superior classification capabilities with remarkable accuracy levels—97.5% in binary categorizations and 94% in multi-class scenarios. The findings from this study underscore the potential of employing advanced Transformer models in the precise classification of breast tumors, thereby contributing to the advancement of diagnostic techniques in oncology.

General Terms
Application of Computer science in Modeling, Image Classification and Deep Learning.

Keywords
Multi-class Classification, Binary Classification, Biomedical Image Processing, Breast Cancer, Benign and Malignant.

1. INTRODUCTION
Breast cancer is currently ranked as the second most common cancer worldwide and continues to be the foremost reason for cancer-induced deaths among women. The importance of early discovery and precise categorization of breast cancer cannot be overstated, as these are critical factors in providing successful treatment and improving the likelihood of positive outcomes for patients.

In recent years, considerable progress has been made in the area of breast cancer classification through the implementation of machine learning (ML) and deep learning (DL) technologies. These strategies bring together a diversity of techniques ranging from established ML algorithms such as Support Vector Machines (SVMs), Random Forests (RFs), and Artificial Neural Networks (ANNs), to more complex DL models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [1]-[3].

The adoption of ML and DL methodologies in the realm of breast cancer evaluation has been transformative, allowing those in the field to derive important conclusions from intricate data. By wielding sophisticated algorithms, these techniques are adept at scrutinizing a multitude of characteristics and configurations within breast cancer imagery or patient records. This leads to a more precise identification of the nature of tumors as either malignant or benign. As a result of integrating ML and DL processes, there has been an uptick in the accuracy of diagnoses, the personalization of treatment plans, and a decrease in unnecessary medical procedures [4]-[7].

In their research, Zhang et al. [8] introduced an innovative CNN-RNN hybrid model tailored for breast cancer type determination. This model harnesses the combined capabilities of CNNs for image-based feature extraction and RNNs for sequential data analysis. Upon evaluation with a collection of breast pathology images, this hybrid model demonstrated superior performance over other leading breast cancer classification models. Similarly, Wang et al. [9] designed a deep learning model utilizing a multi-scale CNN architecture specifically for breast cancer classification. Their model, when assessed using a set of mammogram images, showed enhanced effectiveness in comparison to other advanced deep learning models as well as conventional machine learning techniques.

Beyond the realm of deep learning, traditional machine learning techniques have maintained their relevance in categorizing types of breast cancer. Liu et al. [10] put forward a model using the random forest approach, applying it to breast cancer gene expression data to sort cancer types. Meanwhile, Chen et al. [11] developed a model that relies on a support vector machine algorithm, targeting breast cancer detection through diffusion-weighted MRIs. This SVM-based model was notably accurate, with an 85.5% success rate in differentiating between malignant and benign breast tumours.

The advent of transformer-based models has shown notable promise in breast cancer classification. Originally developed for use in natural language processing, transformers have expanded their reach to include areas such as computer vision and the analysis of medical images. These models are particularly adept at identifying dependencies over large distances and comprehending the contextual nuances within data, attributes that are particularly beneficial for the complex process of medical image examination and feature extraction.

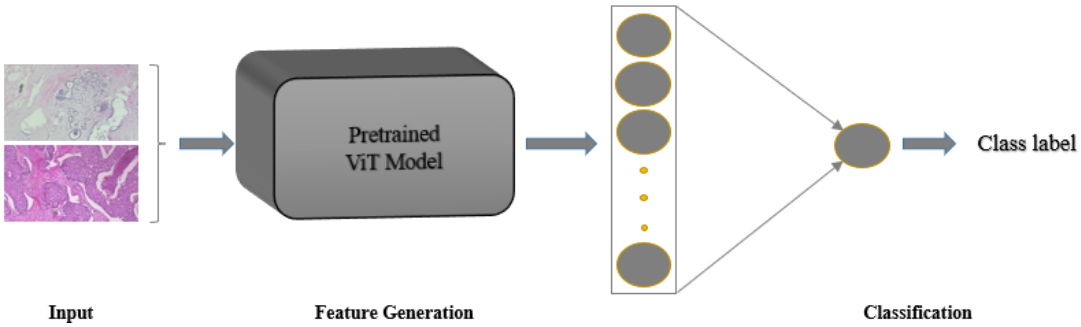


Fig 1: Overview of the Proposed Methodology

Transformers utilize self-attention mechanisms to effectively map out the interplay among different areas of interest in an image. This capacity allows for a more precise and dependable classification of breast cancer. The field has seen increasing exploration into the application of these transformer-based structures, notably Vision Transformers (ViTs), for tasks related to breast cancer categorization. Evidence from these studies points to the enhanced capabilities of transformers to refine the processes of medical image analysis, which could lead to better diagnostic and treatment strategies.

In particular, Xie et al. [12] delved into the deployment of a ViT-based model for the diagnosis of breast cancer using mammography. Their findings highlighted the ViT model's proficiency in grasping the overall context of an image and accurately modeling interactions over long ranges, culminating in bolstered efficiency in the detection and categorization of breast cancer.

In their work, Li et al. [13] examined how combining Vision Transformers (ViTs) with various clinical data, such as histopathological images and patients' clinical histories, could be beneficial. They utilized the distinctive attention mechanism within transformers to selectively focus on pertinent features from both visual and textual information, thereby allowing for a more nuanced and integrated evaluation. Their findings pointed towards the effectiveness of ViTs in handling multiple types of data to improve the accuracy of breast cancer classification and prognostication efforts.

Additionally, Zhang et al. [14] investigated the application of transformers for the segmentation of breast tumors in MRI scans. Their development of a transformer-based model that precisely traced the contours of tumors aids significantly in planning and evaluating treatment options. The study showcased the model's proficiency in understanding spatial and contextual data to achieve accurate tumor segmentation.

Collectively, recent research highlights the promise of transformer-based models, especially Vision Transformers (ViTs), in tasks such as breast cancer classification and segmentation. These advanced models are recognized for their ability to grasp overarching context and to process information over extended sequences, as well as their proficiency in handling data from multiple modes. This is instrumental in offering deeper insights for the improvement of diagnostic precision and the formulation of treatment plans in breast cancer care.

Overall, the literature points to a favourable outlook for both classic machine learning algorithms and deep learning techniques, with a special nod towards deep learning models, in the realm of breast cancer classification. That said, a gap persists that calls for in-depth comparative studies across varied types of breast cancer datasets — including pathology slides, mammographic images, and genetic expression patterns.

Conducting such research is crucial for ascertaining the most effective classification methods for particular contexts within breast cancer diagnostics and treatment, propelling the industry forward and optimizing patient outcomes.

The core contribution of this study is the application of transformer technology to enhance the accuracy of breast cancer categorization. By retraining a model devoted to breast cancer data, this research harnesses the power of the Pretrained Vision Transformer (ViT). The ViT, inherently a deep learning model crafted for image recognition tasks, is a pivotal element in our investigation. Originating from a transformer structure designed for analyzing text-based information, the ViT has been adapted expertly to handle visual data. It approaches image analysis by segmenting images into consistently-sized patches which are then linearly embedded and interpreted by the transformer's mechanisms. In the context of breast cancer classification, our approach is to fine-tune a pretrained ViT model with a specialized dataset of breast cancer visual data. During this fine-tuning phase, the model is trained to identify and learn distinctive visual patterns that signify either the presence or the absence of cancerous conditions. Consequently, it acquires the capability to systematically categorize images into distinct classifications, such as 'benign', 'malignant', or 'normal', based on the visual information it has learned to process.

2. PROPOSED METHOD

This paper focuses on refining a ViT Transformer model for the task of breast cancer classification, as illustrated in Figure 1. The fine-tuning procedure encompasses multiple phases. The initial phase involves data preprocessing, converting raw images into a numerical format compatible with the ViT architecture. During this phase, standardization practices such as normalization are applied, adjusting the pixel intensities of the images to conform to a specific scale.

Let $\boldsymbol{X}$ be the input data, and $\boldsymbol{x}_i$ be the i-th data point in $\boldsymbol{X}$. The normalization process can be represented mathematically as:

$$\boldsymbol{x}_i = normalize(\boldsymbol{X}_i) \quad (1)$$

Let $\boldsymbol{M}$ be the ViT model architecture, which takes as input the normalized image data $\boldsymbol{x}_i$. The ViT model can be mathematically represented as:

$$\boldsymbol{h}_i = \boldsymbol{M}(\boldsymbol{x}_i) \quad (2)$$

where $\boldsymbol{h}_i$ is the output of the ViT model for the i-th input data point.

To fine-tune the ViT model for breast cancer classification, training is conducted using a labeled dataset of breast cancer images. Let $\boldsymbol{Y}$ represent the target labels for the classification task, where $\boldsymbol{y}_i$ denotes the label for the ***i-th*** data point in $\boldsymbol{X}$. The dissimilarity between the predicted output of the ViT model and the true target label is quantified using binary cross-entropy as the chosen loss function. Mathematically, the loss function is represented as follows:

$$L = -(1/\boldsymbol{N}) * sum(\boldsymbol{y}_i * \log(\boldsymbol{p}_i) + \cdots (1 - \boldsymbol{y}_i) * log(1 - \boldsymbol{p}_i)) \quad (3)$$

where $\boldsymbol{N}$ is the total number of data points, $\boldsymbol{p}_i$ is the predicted probability of the positive class for the ***i-th*** data point, and log is the natural logarithm function.

The optimization process involves minimizing the loss function with respect to the model parameters using an optimizer such as stochastic gradient descent (SGD). The optimization process can be mathematically represented as:

$$\boldsymbol{theta} = argmin\_theta\, L(\boldsymbol{theta}) \quad (4)$$

where $\boldsymbol{theta}$ represents the model parameters, and $argmin\_theta$ represents the values of the model parameters that minimize the loss function.

Once the model is trained, we use it to predict the class labels for new, unseen data points. Let $\boldsymbol{x}_{test}$ be the test dataset, and $\boldsymbol{x}_j$ be the j-th data point in $\boldsymbol{x}_{test}$. The predicted probability of the positive class for the j-th data point can be obtained using the trained model as follows:

$$\boldsymbol{p}_j = M(\boldsymbol{x}_j) \quad (5)$$

where $\boldsymbol{p}_j$ is the predicted probability of the positive class for the j-th data point.

3. EXPERIMENTAL RESULTS

The performance assessment of the model utilized the BreakHis breast cancer dataset [16], widely recognized in the field. This dataset comprises 9,109 microscopic images of breast tumor tissue from 82 patients, captured at various magnification levels (40X, 100X, 200X, and 400X). It includes 2,480 benign samples and 5,429 malignant samples, all with dimensions of 700x460 pixels in a 3-channel RGB format and 8-bit depth per channel in PNG format. The benign class encompasses Adenosis (A), Fibroadenoma (F), Tubular Adenoma (TA), and Phyllodes Tumor (PT), while the malignant class includes Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC). Statistical information related to this dataset is provided in Table 1. For visual representation, Figures 2 and 3 display slides of a benign and a malignant breast tumor from the same patient, captured at different magnification levels.

When evaluating performance, common metrics include classification accuracy (Acc), recall (sensitivity), and precision (positive predictivity). Accuracy offers a comprehensive measure of the system's performance across all classes. Insights into the system's overall effectiveness are obtained by analyzing these metrics. They are defined based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) as follows:

Table I. Utilized Breast Cancer Dataset

Classes	Subclasses	Magnification factors				Total
		40x	100x	200x	400x	
Benign	A	114	113	111	106	444
	F	253	260	264	237	1014
	TA	109	121	108	115	453
	PT	149	150	140	130	569
Malignant	DC	864	903	896	788	3451
	LC	156	170	163	137	626
	MC	205	222	196	169	792
	PC	145	142	135	138	560
Total		1995	2081	2013	1820	7909

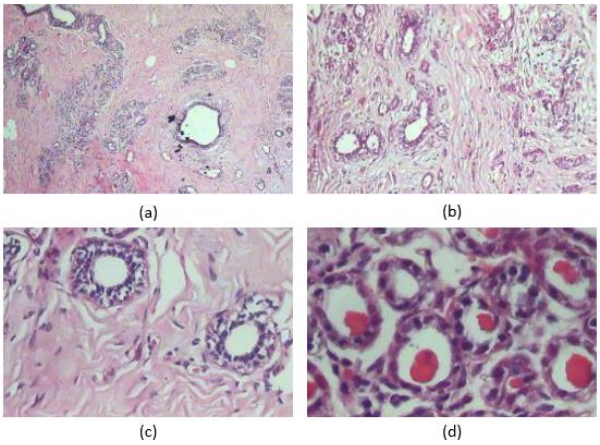


Fig 2: Benign breast tissue samples from the same patient were collected at different levels of magnification: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.

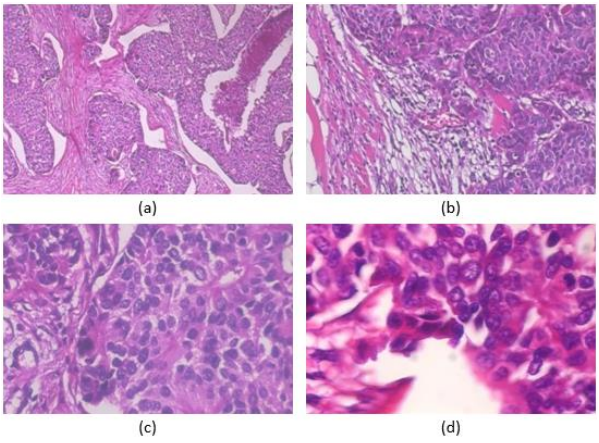


Fig 3: Malignant breast tissue samples from the same patient were collected at different levels of magnification: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

The dataset was partitioned into three distinct subsets to support the model's training, validation, and testing phases. The training set constituted 80% of the data, with the remaining 20% equally divided between validation and test sets, allocating 10% to each. Consequently, the training set included 6,327 samples, whereas both the validation and test sets contained 791 samples each. To ensure result reliability, the experiment was replicated five times, each time training, validating, and testing the model on distinct subsets. The repetition of the experiment enabled the calculation of average classification accuracy over these iterations, offering a thorough assessment of the model's efficacy. This methodical approach, involving multiple experimental runs and average accuracy computation, facilitated the attainment of more dependable and statistically robust outcomes.

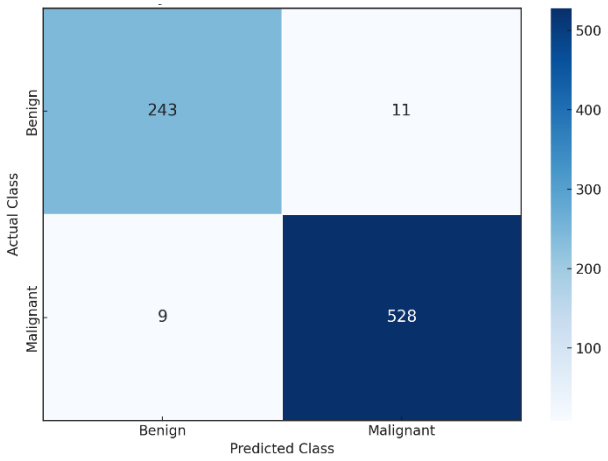
Table II. Comparative Analysis of Various Learning Models and the Proposed Method on the BreakHis Dataset.

Classification Type	Methods	Accuracy
Binary	DensNet201[4]	95.6%
	VGG16 [5]	92.5%
	VGG19 [5]	93.8%
	MobileNet [6]	73.1%
	ResNet-50 [7]	80%
	Proposed	97.5%
Multi-Class	AlexNet [17]	79.85%
	CSDCNN [18]	93.25%
	Proposed	94%

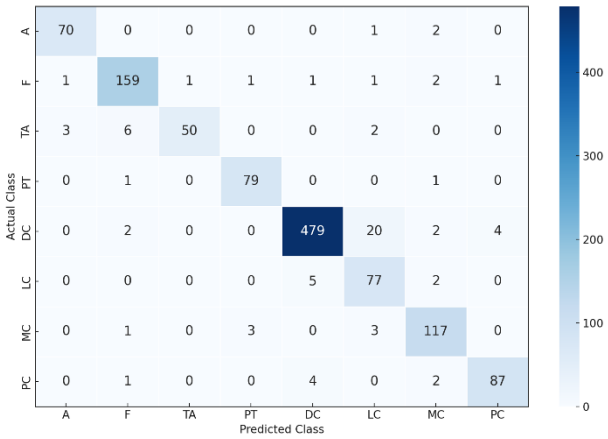
The proposed model's effectiveness is supported by the results presented in Table II, which offer a comparative analysis of its performance when compared to other deep learning architectures. By examining the confusion matrix shown in Figure 4(a), it becomes clear that the model excels in effectively distinguishing between the two classes. Notably, the obtained results demonstrate impressive classification accuracies of 95.7% for the benign class and 98.3% for the malignant class. These high accuracies highlight the model's proficiency in correctly classifying the different types of breast cancer.

Additionally, the model exhibits an impressive precision score of 0.957, indicating its ability to accurately predict malignant tumors approximately 95.7% of the time. This precision metric is highly valuable, especially in the field of medical diagnoses, where precise positive predictions are essential to avoid unnecessary treatments and minimize the occurrence of false positives. Ensuring precise predictions is crucial in minimizing the risk of false alarms and providing patients with the most accurate and appropriate medical advice and interventions.

The model demonstrates an impressive recall of 0.988, reflecting its capability to effectively identify 98.8% of the actual malignant tumor samples. In the context of medical diagnoses, where the consequences of false negatives can be severe, the recall metric holds significant importance. Ensuring high recall helps to minimize the risk of missing positive cases and ensures that patients receive timely and accurate diagnoses for appropriate medical interventions.



(a)



(b)

Fig 4: Confusion Matrices for the Proposed Model. (a) Binary classification. (b) Multi-class classification

Furthermore, the F1 score, which takes into account both precision and recall, is notably high at 0.972. This high F1 score indicates a favourable balance between precision and recall, which is essential for assessing the effectiveness of a classification model. A high F1 score demonstrates that the model performs well in accurately identifying positive cases (precision) while also minimizing the risk of missing positive instances (recall).

Given the provided confusion matrix, it is clear that the classification model exhibits proficient performance in distinguishing between benign and malignant tumors. However, it is crucial to consider the specific problem being investigated, as well as the potential consequences of different error types. In the context of medical diagnoses, false negatives can have severe repercussions, underscoring the importance of optimizing the recall metric. It is essential to prioritize recall to minimize the risk of missing positive cases, which is particularly critical in medical settings where the timely detection of malignant tumors is crucial for effective treatment and patient outcomes. As such, understanding the specific problem domain and the implications of different types of errors is vital for evaluating and improving the performance of a classification model.

Therefore, it is of utmost importance to conduct a thorough evaluation of the performance of a classification model by considering the contextual factors and potential consequences associated with different types of errors. This comprehensive assessment ensures a deeper and more nuanced understanding of the model's effectiveness and greatly assists in making well-informed decisions in real-world applications. By taking into account the specific circumstances and potential implications of the model's performance, we can better assess its suitability and usefulness in practical scenarios.

Upon analyzing the confusion matrix in Figure 4(b), it becomes evident that the model exhibits a noteworthy level of accuracy when it comes to classifying cases of benign breast cancer. Specifically, it demonstrates a high degree of precision in predicting adenosis and fibroadenoma, with a remarkable success rate of 70 and 159 samples respectively. What is particularly noteworthy is that the model does not misclassify any samples for these benign classifications, which suggests its ability to accurately differentiate them from other forms of breast cancer. These findings provide strong evidence indicating that the model possesses a robust ability to distinguish cases of benign breast cancer within the BreakHis dataset.

However, the model encounters challenges in classifying phyllodes tumors. Although it correctly identifies 50 cases as true positives, indicating its ability to recognize phyllodes tumours, there are instances where it misclassifies samples as other types of cancer and incorrectly predicts other classes as phyllodes tumours. This suggests that the model may have difficulty accurately distinguishing phyllodes tumours from other forms of breast cancer. Further investigation is needed to understand and potentially refine the model's performance in accurately classifying phyllodes tumours.

Turning our attention to malignant breast cancer cases, the model showcases strong performance in classifying tubular adenoma. It achieves an impressive true positive count of 79, highlighting its proficiency in accurately identifying this specific type of malignant breast cancer. Furthermore, the occurrence of only one false positive indicates a high level of specificity in the classification of tubular adenoma.

The model's remarkable performance in classifying ductal carcinoma is evident from its true positive count of 479, indicating its exceptional ability to correctly identify a substantial number of cases. However, the presence of a few misclassification errors, where samples from other classes are incorrectly predicted as ductal carcinoma, suggests a potential limitation in the model's ability to discriminate ductal carcinoma from other breast cancer subtypes. This highlights the need for further refinement and improvement in the model's discriminatory capabilities, particularly for ductal carcinoma, within the BreakHis dataset.

When it comes to lobular carcinoma, the model shows great accuracy in predicting this specific form of malignant breast cancer, with a true positive count of 77. However, there are cases where samples from different classes are incorrectly

classified as lobular carcinoma, indicating the presence of misclassification errors. This emphasizes the need to improve the model's precision in distinguishing lobular carcinoma from other types of breast cancer within the BreakHis dataset.

The model's proficiency in classifying mucinous carcinoma is evident from its true positive count of 117, demonstrating its capability to correctly identify a significant number of cases. However, the presence of three false positives indicates a potential limitation in the model's ability to accurately distinguish mucinous carcinoma from other breast cancer subtypes. This highlights the need for further refinement and improvement in the model's specificity and accuracy, particularly for mucinous carcinoma.

The model's performance in classifying papillary carcinoma is noteworthy, with a true positive count of 87, indicating its ability to correctly identify a substantial number of papillary carcinoma cases. However, the presence of four false positives suggests a potential limitation in the model's ability to distinguish papillary carcinoma from other breast cancer subtypes. This highlights the need for further refinement and improvement in the model's precision and discrimination capabilities, particularly for papillary carcinoma, within the BreakHis dataset.

In summary, the provided confusion matrix offers crucial insights into the model's classification performance for different types of breast cancer in the BreakHis dataset. It demonstrates the model's strengths in accurately predicting benign cases, while also highlighting areas that require further investigation and potential refinement, particularly in distinguishing certain malignant types from other classes. These findings contribute to the advancement of research in breast cancer detection and provide valuable guidance for optimizing the model's diagnostic capabilities in the BreakHis dataset.

4. CONCLUSIONS

This research introduces a novel deep learning approach utilizing transformers for breast cancer image classification, with a methodology that encompasses feature extraction and classification stages. Extensive experimentation on a benchmark dataset has validated the efficacy and established the superiority of this approach. By integrating transformers, this work sets a new benchmark in the field and signals an innovative direction for future studies. Further developments could include refining transformer architectures, enhancing explainability for clinical use, and expanding dataset diversity. Such advances have the potential to significantly contribute to the precision of breast cancer diagnostics and prognostics, ultimately improving patient outcomes.

5. ACKNOWLEDGMENTS

The author would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding this Research Group No. (RG-1438-071).

6. REFERENCES

[1] Md. Mahbubur Rahman, Md. Saikat Islam Khan, Hafiz Md. Hasan Babu, "BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer," Array, vol. 16, pp. 100256, 2022.

[2] B. Xu et al., "Attention by Selection: A Deep Selective Attention Approach to Breast Cancer Classification," in IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1930-1941, June 2020, doi: 10.1109/TMI.2019.2962013.

[3] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen and Y. -D. Zhang, "ResNet-SCDA-50 for Breast Abnormality Classification," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 1, pp. 94-102, 1 Jan.-Feb. 2021, doi: 10.1109/TCBB.2020.2986544.

[4] M. Huang, S. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700-4708. doi: 10.1109/CVPR.2017.243.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proceedings of the International Conference on Learning Representations, 2015. arXiv: 1409.1556.

[6] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778. doi: 10.1109/CVPR.2016.90.

[8] Zhang, H., Lu, S., Ren, J., and Liu, J., "A hybrid CNN-RNN model for breast cancer classification," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 3, pp. 2817-2828, 2021.

[9] Wang, X., Sun, Y., and Li, B., "Multi-scale convolutional neural network for breast cancer classification in mammography images," Journal of Healthcare Engineering, vol. 11, no. 3, pp. 401-409, 2020.

[10] Liu, X., Huang, Y., Wang, C., and Zeng, X., "A random forest-based breast cancer classification model using gene expression profiles," Journal of Healthcare Engineering, vol. 11, no. 2, pp. 235-244, 2020.

[11] Chen, X., Li, Y., and Li, J., "A Support Vector Machine-Based Model for Breast Cancer Classification Using Diffusion-Weighted Magnetic Resonance Imaging," Journal of Healthcare Engineering, vol. 12, no. 3, pp. 2817-2828, 2021.

[12] J. Xie, et al., "Breast cancer diagnosis based on vision transformer," International Journal of Advanced Science and Technology, vol. 30, no. 5, pp. 2826-2836, 2021.

[13] Z. Li, et al., "Integrating Vision Transformer with clinical data for breast cancer classification," Frontiers in Medicine, vol. 9, p. 841457, 2022.

[14] Y. Zhang, et al., "Transformer-based breast tumor segmentation in MRI scans," Medical Image Analysis, vol. 78, p. 102183, 2023.

[15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ..., and Hinton, G., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3156-3164, 2021.

[16] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A Dataset for Breast Cancer Histopathological Image Classification,” IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[17] F.A. Spanhol, L.S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in International Joint Conference on Neural Networks, 2016, pp. 2560-2567.

[18] Z. Han, B. Wei, and Y. Zheng, "Breast cancer multi-classification from histopathological images with structured deep learning model," Scientific Reports, vol. 7, no. 1, p. 4172, 2017.