David vs. Goliath: Large Foundation Models are not Outperforming Small Models in Multi-view Mammogram Breast Cancer Prediction

Xuxin Chen[a,b], Mingzhe Hu[a], Ke Zhang[b,c], Neman Abdoli[b], Youkabed Sadri[b], Patrik Gilley[b], Omkar Saiswaroop Varma Chekuri[d], Farid H. Omoumi[e], Yuchen Qiu[b], Bin Zheng[b], Xiaofeng Yang*[a]

[a]Radiation Oncology and Winship Cancer Institute, Emory School of Medicine, Atlanta, GA 30322
[b]School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019
[c]Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK 73019
[d]School of Computer Science, University of Oklahoma, Norman, OK 73019
[e]School of Engineering Medicine, Texas A&M University, Houston, TX 77030
*Correspondence: xiaofeng.yang@emory.edu

ABSTRACT

Recent *Foundation Models* have begun to yield remarkable successes across various downstream medical imaging applications. Yet, their potential within the context of multi-view medical image analysis remains largely unexplored. This research aims to investigate the feasibility of leveraging foundation models for predicting breast cancer from multi-view mammograms through parameter-efficient transfer learning (PETL). PETL was implemented by inserting lightweight *adapter* modules into existing pre-trained transformer models. During model training, the parameters of the adapters were updated while the pre-trained weights of the foundation model remained fixed. To assess the model's performance, we retrospectively assembled a dataset of 949 patients, with 470 malignant cases and 479 normal or benign cases. Each patient has four mammograms obtained from two views (CC/MLO) of both the right and left breasts. The large foundation model with 328 million (M) parameters, finetuned with adapters comprising only 3.2M tunable parameters (about 1% of the total model parameters), achieved a classification accuracy of 78.9% ± 1.7%. This performance was competitive but slightly inferior to a smaller model with 36M parameters, finetuned using traditional methods, which attained an accuracy of 80.4% ± 0.9%. The results suggest that while foundation models possess considerable potential, their efficacy in medium-sized datasets and in transitioning from single-view to multi-view image analysis, particularly where reasoning feature relationships across different mammographic views is crucial, can be challenging. This underscores the need for innovative transfer learning approaches to better adapt and generalize foundation models for the complex requirements of multi-view medical image analysis.

Keywords: Foundation model; transformer; parameter-efficient transfer learning; adapter; multi-view image; mammogram; breast cancer

1. INTRODUCTION

The "pretraining then finetuning" paradigm[1] has played an important role in the field of medical imaging. While this approach has garnered wide attention since the early days of deep learning applications[2, 3], its significance has surged in very recent years with the rise of Vision Transformer (ViT)[4, 5]. Notably, an increasing emphasis on both the size of the models and the scale of the training data has led to the rise of large-scale, pretrained models, often referred to as *Foundation Models*[6], such as *Segment Anything Model* (SAM)[7]. Despite the promise of these large models, their direct application to medical imaging tasks, especially those requiring domain-specific knowledge, has often resulted in performances that do not match up to those achieved by state-of-the-art (SOTA), smaller deep learning models[8-10]. This suggests that the larger scale of these models does not necessarily translate to an immediate performance boost in specialized medical imaging tasks.

Medical Imaging 2024: Computer-Aided Diagnosis, edited by Weijie Chen, Susan M. Astley, Proc. of SPIE Vol. 12927, 129270X · © 2024
SPIE · 1605-7422 · doi: 10.1117/12.3007054

In response to this, the concept of parameter-efficient transfer learning (PETL), which first emerged from natural language processing (NLP)[11, 12], has become a promising research area. At the core of PETL is inserting lightweight adaptation modules, such as adapters or prompt tokens, into pretrained foundation models. This technique aims to exclusively finetune the parameters of these adaptation modules while keeping the pre-trained weights frozen[13] to attain strong transfer learning performance. This offers a powerful alternative to fully finetuning entire large models. It strikes a balance between model adaptability and efficiency, particularly effective for tasks with small to medium-sized datasets. For instance, a recent study proposed inserting adapters into ViT blocks of the SAM model, achieving SOTA performance in medical image segmentation tasks[14]. It is worth noting, however, that existing works have concentrated their efforts on applying PETL to tailor foundational models for the analysis of single-view medical images. In contrast, the exploration of these techniques in multi-view medical images, which are common in clinical practice, remains largely untapped.

Multi-view medical image analysis refers to analyzing and interpreting medical images from multiple viewpoints to extract meaningful information about a patient's anatomy or pathology[15]. Combining complementary information from multiple views can lead to more comprehensive insights that might not be apparent from a single imaging view alone[16]. In the context of breast cancer screening, for instance, the nuanced reading and interpretation of mammograms are crucial. For a typical mammography screening, each patient has four mammograms acquired from two views of two sides, namely, craniocaudal (CC) and mediolateral oblique (MLO) view of the left (L) and right (R) breasts. These four images are distinct yet complementary, contributing to a holistic understanding of subtle irregularities, such as tumors, calcifications, and indicative markers of breast cancer. Radiologists heavily rely on such multi-view knowledge, looking for correspondence between features in different views of the same breast (ipsilateral comparison) and differences in the same view between the two breasts (bilateral comparison)[17-20]. This comprehensive analysis is critical to a more precise and informed diagnosis. In this domain, smaller models have shown commendable performance, often surpassing larger models[15, 19]. This effectiveness of smaller models in multi-view mammography, a field characterized by more limited and specialized datasets, raises important questions about the advantages and limitations of both large and small models.

Our study aims to investigate the potential of adapting *Goliaths*—large foundation models for breast cancer prediction from multi-view mammograms, alongside a comparison with the nimble *Davids*—smaller models. Interestingly, we observed that fully finetuning large foundation models led to suboptimal results compared to smaller models. This suggests that foundation models' larger scale and complexity may not always be advantageous, especially in contexts where domain-specific knowledge and dataset size are crucial factors. However, the use of adapters in large models did show improved efficacy, indicating the potential of PETL techniques in enhancing the performance of these models for multi-view mammogram analysis. This exploration contributes to a deeper understanding of the role of model size and adaptation techniques in multi-view medical imaging. It highlights the ongoing relevance and advantages of both large and small models in the field.

2. MATERIALS AND METHODS

2.1 Dataset

From an established full-field digital mammography (FFDM) image database[21], we retrospectively constructed a dataset encompassing 3,796 mammograms from 949 patients for this study. Notably, each patient has four distinct mammograms from both the CC and MLO views of the left and right breasts, denoted as LCC, RCC, LMLO, and RMLO, respectively. All mammograms were captured using Hologic Selenia digital mammography equipment, precisely calibrated with a fixed pixel dimension of 70μm. The dimensions of the acquired mammograms vary depending upon the individual breast size, resulting in two potential sizes: either 2558×3327 or 3327×4091 pixels. For analytical purposes, the original FFDM images underwent subsampling via a pixel averaging method, incorporating a kernel size of 5×5 pixels. As a result, the dimensions of the two original FFDM image variants were effectively reduced to 512×666 and 666×819 pixels, respectively, while maintaining a pixel size of 0.35mm.

A subset of 349 cases was conclusively assessed by radiologists as screening negative or benign (BIRAD 1 or 2). The remaining cases were recalled and recommended for biopsy due to the identification of potentially concerning soft-tissue masses. After histopathological analysis of the biopsy specimens, 130 and 470 cases were definitively confirmed as

benign and malignant, respectively. To facilitate binary classification, we grouped all cases classified as screening negative or benign lesions into a unified benign class, distinctly differentiating these from cases under the categorization of cancer or malignant cases. In other words, our study focused exclusively on predicting the likelihood of malignancy within mammographic cases.

2.2 ViT-based framework for multi-view mammography

Figure 1 provides an overview of the ViT-based framework for multi-view mammogram analysis. Mammograms (LCC, RCC, LMLO, RMLO) are transformed into patch embeddings with positional embeddings. The CC and MLO view embeddings are sent into view-specific ViT backbones. Each backbone consists of N transformer blocks, divided into N_l local and N_g global blocks. The local blocks focus on modeling local feature relations within each individual view of the mammogram. Local transformer outputs are concatenated into a sequence, passing through the global blocks to learn long-range, cross-view features (i.e., bilateral feature difference and ipsilateral feature correspondence). The last global block's class token enters a multilayer perceptron (MLP) head for binary classification.

Our backbone is a hybrid, ImageNet-pretrained ViT-large model with 24 transformer blocks, enhanced with a 50-layer ResNet for patch embedding[4, 22]. This hybrid model combines traditional convolutional neural networks (CNNs) at early stages, thereby improving its ability to detect and learn local patterns missed by pure transformers and improving both its generalizability and trainability[23-25]. As a comparative measure, we also explore a smaller hybrid ViT model with 12 transformer blocks, which is also ImageNet-pretrained but incorporates a 26-layer ResNet for patch embedding. In our study, these are referred to as ViT-Large-R50 (328M parameters) and ViT-Small-R26 (36M parameters).

Consistent with previous studies[19], we have made the local blocks of the CC and MLO views share weights. This is beneficial in streamlining the training process by employing a single hybrid ViT model for multi-view analysis, significantly reducing computational time. Previous findings[19] suggest that the number of local and global blocks in a pure vision transformer critically influences classification performance in multi-view mammogram analysis. Accordingly, we investigated the effects of varying these blocks in our hybrid ViT model.

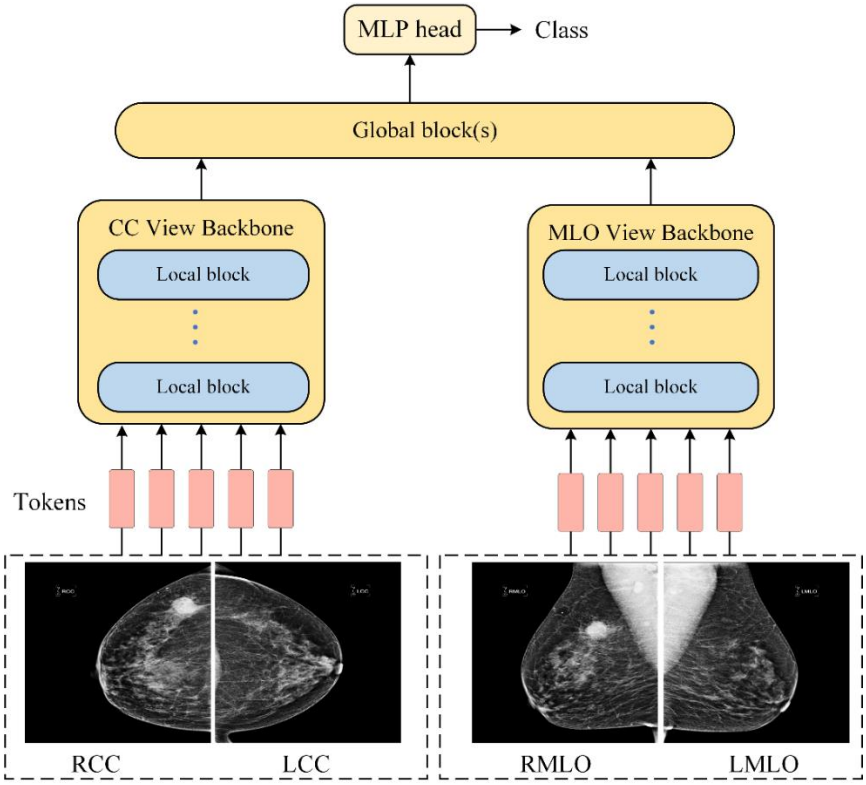

Figure 1. Overview of the ViT-based framework for multi-view mammogram analysis.

2.3 Multi-view ViT finetuning with Adapters

A ViT consists of multiple transformer blocks, each featuring multi-head self-attention (MSA) and a multilayer perceptron (MLP) layer, as illustrated in Figure 2a. The Adapter module, depicted in Figure 2b, employs a bottleneck design with a central activation layer (e.g., GELU) sandwiched between two fully connected (FC) layers. The first FC layer reduces the feature dimension by a bottleneck ratio and the second restores it to the original dimension. A larger bottleneck ratio means fewer tunable parameters during model finetuning, affecting performance.

Adapter insertion in a ViT block can vary: it can be added to just the MSA layer, the MLP layer, or both simultaneously. Research shows that inserting adapters in both MSA and MLP layers typically enhances the model's adaptability, a practice we adopt in this study. In addition, adapters can be inserted sequentially (Figure 2c) before/after or in parallel with the MSA and MLP. We will explore the effects of these different adapter configurations, including the bottleneck ratio and the insertion method.

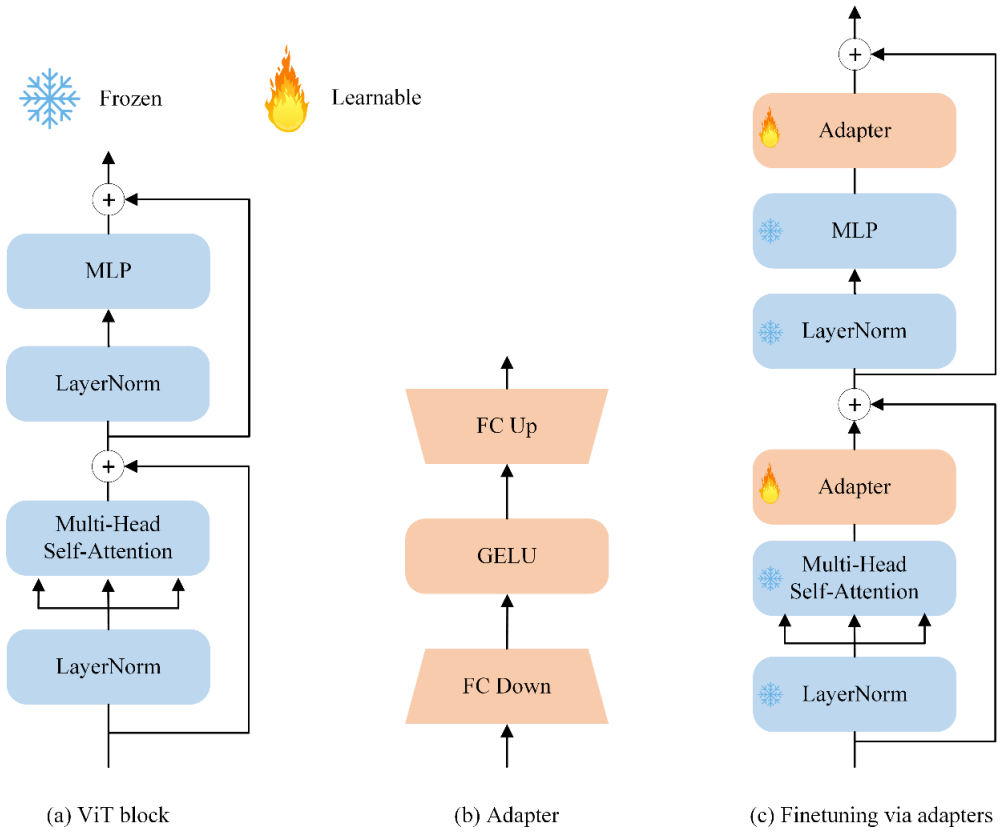

Figure 2. (a) A standard ViT block. (b) An Adapter module. (c) Adapters for ViT finetuning.

When fine-tuning large ViT models, only the adapters' weights were updated, keeping the ViT block weights frozen. In our study, the small model was fully finetuned using the mammogram dataset without adapters, while the large model employed both adapter-based and traditional fully finetuning methods. We largely follow the training setting in Touvron et al. (2021). Specifically, the model was finetuned 400 epochs using AdamW[22] optimizer with a batch size 16. The base learning rate is 5e-4. After warming up in the first 10 epochs, the learning rate decays following a cosine schedule. For data augmentation, we employed normalization, random horizontal flips, and random erasing. Given that the original mammograms are 12-bit images with 4096 grayscale levels, we adapted them for the pre-trained Transformer by expanding across RGB channels and normalizing pixel values to the [0, 255] range, resizing all images to 224×224 pixels. To evaluate performance and generalization, we used a five-fold cross-validation technique, computing mean accuracy and standard deviation. All experiments were carried out in PyTorch on a single NVIDIA Tesla V100 GPU with 32GB VRAM.

3. RESULTS

Table 1 summarizes the classification performance of both small and large ViTs on the multi-view mammogram dataset. The small hybrid ViT (36M tunable parameters) achieves a classification accuracy of 80.4% ± 0.9% through traditional finetuning, as evaluated by the five-fold cross-validation process. In comparison, the large hybrid ViT (328M tunable parameters) initially exhibits a much lower classification accuracy of 75.6% ± 1.7% with traditional finetuning. Notably, upon finetuning with adapters (3.2M tunable parameters), the large model's performance jumps to 78.9% ± 1.7%, marking a 4.4% relative improvement. We also observed that the large model's performance tends to decline with an increase in local blocks and a corresponding decrease in global blocks.

In Table 2, we investigate the influence of different bottleneck ratios on the efficacy of adapters in model performance. Among the evaluated ratios (16, 32, 64), the ratio of 32 emerges as the most effective, suggesting that extremely high or low bottleneck ratios leads to suboptimal adapter finetuning performance.

Table 3 compares the performance of different adapter configurations, specifically assessing sequential versus parallel insertion methods and their respective locations. The results show that the sequential insertion of adapters before the MSA and MLP layers significantly weakens the adaptability and generalizability of the large model on the multi-view mammogram dataset. The optimal performance is achieved through the sequential insertion of adapters after the MSA and MLP layers, which marginally surpasses parallel adapters.

Table 1. Summary of the case classification accuracy (ACC) generated by the models, accompanied by the corresponding standard deviation (STD). "Param" represents parameters.

Model	Finetuning Method	Param (M)	Tunable Param (M)	N_l	N_g	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ACC (%) ± STD
ViT-Small-R26	Full finetuning	36	36	0	12	80.0	80.5	79.5	82.1	79.9	**80.4 ± 0.9**
ViT-Large-R50	Linear probing	328	0.004	0	24	76.3	76.3	77.4	75.3	72.5	75.6 ± 1.7
	Full finetuning	328	328	0	24	80.0	75.3	72.1	75.8	79.4	76.5 ± 2.9
	Adapter	331	3.2	0	24	81.1	80.5	77.4	78.9	76.7	**78.9 ± 1.7**
	Adapter	331	3.2	4	20	82.1	76.3	78.4	81.1	75.1	78.6 ± 2.7
	Adapter	331	3.2	8	16	81.1	75.8	78.4	77.4	75.7	77.7 ± 2.0
	Adapter	331	3.2	16	8	80.0	74.7	82.6	77.4	76.2	78.2 ± 2.8

Table 2. Effect of bottleneck ratio of Adapters.

Bottleneck Ratio	Param (M)	Tunable Param (M)	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ACC (%) ± STD
64	330	1.6	82.1	75.3	77.9	80.5	75.1	78.2 ± 2.8
32	331	3.2	81.1	80.5	77.4	78.9	76.7	**78.9 ± 1.7**
16	334	6.3	82.1	77.9	80.0	78.9	75.7	78.9 ± 2.1

Table 3. Effect of sequential vs. parallel designs of Adapters and insert location.

Insert Form	Insert Location	Tunable Param (M)	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ACC (%) ± STD
Sequential	Before MSA &MLP	3.2	75.3	73.2	76.3	75.3	75.1	75.0 ± 1.0
Sequential	After MSA &MLP	3.2	81.1	80.5	77.4	78.9	76.7	**78.9 ± 1.7**
Parallel	Parallel to MSA&MLP	3.2	80.0	77.9	81.1	76.8	77.8	78.7 ± 1.6

4. DISCUSSION AND CONCLUSION

Large foundation models are a rapidly evolving domain characterized by their potential for solid generalization and adaptability across various downstream tasks. This study focused on their application in multi-view medical images, specifically in multi-view mammogram analysis for breast cancer prediction. Our investigation generates a more nuanced picture of their effectiveness in the context of mammography.

We conducted a comparative analysis between a large hybrid ViT model and a smaller counterpart. Interestingly, the larger model, with a classification accuracy of 78.9% ± 1.7%, was outperformed by the smaller model, which achieved an accuracy of 80.4% ± 0.9%. This challenges the prevailing belief that larger models outperform smaller ones in downstream tasks, particularly in specialized fields such as multi-view mammogram analysis. The reasons for this underperformance seem at least twofold. Firstly, it may relate to the interplay between data availability and model capacity. In scenarios with limited data, large models have shown remarkable performance in few-shot learning[14, 26]. However, our study utilized a medium-sized multi-view mammogram dataset. In such a context, where the data size is sufficient to fully optimize a smaller model but not large enough for a large foundation model, the benefits of using a large model become less evident. Nonetheless, the use of adapter modules with 3.2M tunable parameters, which is equal to finetuning about only 1% of the total parameters of the large model, achieved performance on par with the high-performing small model. This demonstrates the potential of adapters to achieve efficient parameter utilization in transfer learning, particularly beneficial in multi-view mammogram analysis.

The second aspect concerns the domain-specific characteristics of multi-view mammography, which hinges on effectively reasoning feature relations across different mammographic views. Previous studies[15, 19] have shown that lightweight vision transformers or transformer blocks can perform well in breast cancer prediction from multi-view mammograms. This is likely because smaller models, with fewer parameters and medium-sized datasets, can more readily adapt from single-view feature learning to multi-view reasoning. In contrast, most publicly available large-scale ViTs, exclusively trained on single-view images like ImageNet[27] and JFT-300M[28], although possessing more robust feature representations, may face challenges in swiftly transitioning from single-view to multi-view image analysis. Thus, the somewhat subpar performance of large foundation models is not necessarily a reflection of poor feature learning but rather a lack of effective techniques to correlate and reason these single-view features within a multi-view framework. This leads us to speculate on the necessity for developing methods that can adeptly integrate the single-view features of large foundation models for multi-view medical imaging applications.

We also found that increasing local transformer blocks while reducing global ones led to poorer performance. This aligns with insights from previous research[19], which suggested a limited necessity for local transformer blocks in the context of multi-view mammogram analysis. Due to the effectiveness of convolutional approaches for patch embeddings in capturing local features/patterns, using local transformer blocks to perform a similar function became unnecessary, maybe even redundant. The critical aspect, therefore, was to leverage the inherent strength of transformers in capturing long-range dependencies across different views. This finding underscores the importance of optimizing the balance between local and global transformer blocks, focusing on harnessing the capabilities of global blocks for enhanced performance in multi-view mammogram analysis.

Despite these insights, we recognize this study has several limitations. Firstly, for simplicity, the mammograms were resized to 224×224 from their original high-resolution format. While common in mammogram analysis to fit the input size of pre-trained models[18, 29], this practice might omit critical details best captured at the original resolution. Previous studies[30, 31] have shown that optimal performance in mammogram analysis is often achieved with the original resolution. Therefore, we believe future research adapting large foundation models to high-resolution mammogram analysis would be highly valuable. Secondly, our exploration was limited to a single type of adapter due to time constraints. Other emerging forms, such as LoRA[12], SSF[32], etc., might offer better adaptability for large foundation models and need further investigation. In addition, our dataset, sourced from a single cohort, might not fully represent the diversity required for comprehensive model evaluation. Future studies should aim to include more diversified datasets from various institutions and experiment with varying sizes of training data. This would enable a more thorough comparison of the performance of large and small models under different conditions, providing deeper insights into the challenges and benefits of using large foundation models in multi-view medical imaging tasks.

In conclusion, this study leveraged large foundation models for breast cancer prediction in multi-view mammogram analysis through adapter finetuning. However, these models demonstrated some underperformance compared to their smaller counterparts in terms of accuracy. This suggests that while large foundation models hold significant potential in medical imaging, there is a crucial need for more effective strategies to adapt these models for multi-view feature fusion in medical images. The exploration of new adapter types and the application of these models to high-resolution and diverse datasets could pave the way for realizing their full potential in this domain.

ACKNOWLEDGEMENTS

This research was funded in part by National Institutes of Health, USA, under grant number P20 GM135009 and R01CA272991.

REFERENCES

[1] T. Yang, Y. Zhu, Y. Xie *et al.*, "Aim: Adapting image models for efficient video action recognition," arXiv preprint arXiv:2302.03024, (2023).
[2] H.-C. Shin, H. R. Roth, M. Gao *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging, 35(5), 1285-1298 (2016).
[3] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," IEEE transactions on medical imaging, 35(5), 1299-1312 (2016).
[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, (2020).
[5] X. Chen, X. Wang, K. Zhang *et al.*, "Recent advances and clinical applications of deep learning in medical image analysis," Medical Image Analysis, 102444 (2022).
[6] R. Bommasani, D. A. Hudson, E. Adeli *et al.*, "On the opportunities and risks of foundation models," arXiv preprint arXiv:2108.07258, (2021).
[7] A. Kirillov, E. Mintun, N. Ravi *et al.*, "Segment anything," arXiv preprint arXiv:2304.02643, (2023).
[8] R. Deng, C. Cui, Q. Liu *et al.*, "Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging," arXiv preprint arXiv:2304.04155, (2023).
[9] S. He, R. Bao, J. Li *et al.*, "Accuracy of segment-anything model (sam) in medical image segmentation tasks," arXiv preprint arXiv:2304.09324, (2023).
[10] T. Wald, S. Roy, G. Koehler *et al.*, "SAM. MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model."
[11] N. Houlsby, A. Giurgiu, S. Jastrzebski *et al.*, "Parameter-efficient transfer learning for NLP." 2790-2799.
[12] E. J. Hu, Y. Shen, P. Wallis *et al.*, "Lora: Low-rank adaptation of large language models," arXiv preprint arXiv:2106.09685, (2021).
[13] S. Jie, and Z.-H. Deng, "Convolutional bypasses are better vision transformer adapters," arXiv preprint arXiv:2207.07039, (2022).
[14] J. Wu, R. Fu, H. Fang *et al.*, "Medical sam adapter: Adapting segment anything model for medical image segmentation," arXiv preprint arXiv:2304.12620, (2023).
[15] G. v. Tulder, Y. Tong, and E. Marchiori, "Multi-view analysis of unregistered medical images using cross-view transformers." 104-113.
[16] X. Wu, H. Hui, M. Niu *et al.*, "Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study," European Journal of Radiology, 128, 109041 (2020).
[17] Z. Yang, Z. Cao, Y. Zhang *et al.*, "MommiNet-v2: Mammographic multi-view mass identification networks," Medical Image Analysis, 73, 102204 (2021).
[18] G. Carneiro, J. Nascimento, and A. P. Bradley, "Automated analysis of unregistered multi-view mammograms with deep learning," IEEE transactions on medical imaging, 36(11), 2355-2365 (2017).

[19] X. Chen, K. Zhang, N. Abdoli *et al.*, “Transformers Improve Breast Cancer Diagnosis from Unregistered Multi-View Mammograms,” Diagnostics, 12(7), 1549 (2022).
[20] Y. Liu, F. Zhang, C. Chen *et al.*, “Act like a radiologist: towards reliable multi-view correspondence reasoning for mammogram mass detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), 5947-5961 (2021).
[21] B. Zheng, J. H. Sumkin, M. L. Zuley *et al.*, “Computer-aided detection of breast masses depicted on full-field digital mammograms: a performance assessment,” The British Journal of Radiology, 85(1014), e153-e161 (2012).
[22] H. Touvron, M. Cord, M. Douze *et al.*, "Training data-efficient image transformers & distillation through attention." 10347-10357.
[23] Z. Dai, H. Liu, Q. V. Le *et al.*, “Coatnet: Marrying convolution and attention for all data sizes,” Advances in Neural Information Processing Systems, 34, 3965-3977 (2021).
[24] T. Xiao, M. Singh, E. Mintun *et al.*, “Early convolutions help transformers see better,” Advances in neural information processing systems, 34, 30392-30400 (2021).
[25] Z. Tu, H. Talebi, H. Zhang *et al.*, "Maxvit: Multi-axis vision transformer." 459-479.
[26] P. Gao, S. Geng, R. Zhang *et al.*, “Clip-adapter: Better vision-language models with feature adapters,” International Journal of Computer Vision, 1-15 (2023).
[27] J. Deng, W. Dong, R. Socher *et al.*, "Imagenet: A large-scale hierarchical image database." 248-255.
[28] C. Sun, A. Shrivastava, S. Singh *et al.*, "Revisiting unreasonable effectiveness of data in deep learning era." 843-852.
[29] G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models." 652-660.
[30] K. J. Geras, S. Wolfson, Y. Shen *et al.*, “High-resolution breast cancer screening with multi-view deep convolutional neural networks,” arXiv preprint arXiv:1703.07047, (2017).
[31] D. Ribli, A. Horváth, Z. Unger *et al.*, “Detecting and classifying lesions in mammograms with deep learning,” Scientific reports, 8(1), 4165 (2018).
[32] D. Lian, D. Zhou, J. Feng *et al.*, “Scaling & shifting your features: A new baseline for efficient model tuning,” Advances in Neural Information Processing Systems, 35, 109-123 (2022).