TURKISH JOURNAL OF SCIENCE
VOLUME 10, ISSUE 2, 95–104
ISSN: 2587–0971

https://www.tjoscience.com

A Comparative Performance Analysis of Vision Transformer Architectures for Breast Ultrasound Image Classification

Yigitcan CAKMAK[a], Ishak PACAL[a]

[a]*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye.*

Abstract. Although ultrasonography plays a critical role in the early detection of breast cancer, its limitations, such as operator dependency, necessitate the development of objective analysis methods. In response to this need, deep learning models based on the Vision Transformer (ViT) architecture present promising solutions. This investigation comparatively assesses the performance of four modern Transformer architectures Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base for the classification of breast ultrasound images into benign, malignant, and normal categories. Conducted on the publicly available "Breast Ultrasound Images Dataset," the study integrated dynamic data augmentation techniques to enhance model generalization. The empirical results demonstrated a statistically significant superiority of the DeiT-Base model, which achieved 94.30% accuracy and a 93.85% F1-score. While ViT-Base and Swin-Base delivered competitive outcomes, BEiT-Base exhibited the lowest performance with 66.46% accuracy. These findings indicate that for the analysis of limited and distinct datasets such as breast ultrasound, data-efficient training strategies like the knowledge distillation employed by DeiT may be more impactful than architectural differences alone. Moreover, these approaches hold considerable potential for future integration into clinical decision support systems.

1. Introduction

Breast cancer, a malignant tumor characterized by the uncontrolled proliferation of epithelial cells within the breast tissue, represents a significant global public health issue. Its pathogenesis is attributed to an intricate interplay of genetic, hormonal, lifestyle, and environmental factors [1–3]. Within the management of this disease, early diagnosis is the paramount factor directly influencing treatment success and patient survival; consequently, non-invasive diagnostic assessments emerge as a cornerstone strategy [4–6]. Mammography, recognized as the gold standard for population-based screening programs [7, 8], possesses a notable limitation: its diagnostic sensitivity is diminished, particularly in breast structures with dense fibroglandular tissue and in younger women. This circumstance gives rise to a scientific and clinical imperative for complementary imaging modalities capable of addressing these diagnostic voids [9].

In response to this requirement, ultrasonography stands as a pivotal complementary method due to advantages such as the absence of ionizing radiation, low cost, and the provision for dynamic evaluation [10]. It is an invaluable tool for clarifying mammographic findings, differentiating between cystic and solid

Corresponding author: YC mail address: ygtcncakmak@gmail.com ORCID:0009-0008-7227-9182, IP ORCID:0000-0001-6670-2169
Received: 21 June 2025; Accepted:19 July 2025; Published: 30 September 2025
Keywords. Breast cancer, Deep learning, Medical imaging
2010 *Mathematics Subject Classification*. 92C50, 68T07, 92C55.
Cited this article as: Cakmak, Y., & Pacal, I. (2025). A Comparative Performance Analysis of Vision Transformer Architectures for Breast Ultrasound Image Classification. Turkish Journal of Science, 10(2), 95-104.

masses, and guiding interventional procedures like biopsies. It significantly enhances diagnostic performance in dense breast tissue by rendering lesions more conspicuous [11]. Nevertheless, the effectiveness of ultrasonography exhibits a high degree of dependence on operator experience and interpretation, a situation that can lead to diagnostic inconsistencies among different specialists. This limitation, combined with its inadequacy in detecting microcalcifications, underscores the necessity for objective, reproducible analysis methods to enhance standardization and mitigate subjectivity in the interpretation of ultrasound images [12, 13].

In recent years, Vision Transformer (ViT) architectures are spearheading revolutionary advancements in the field of medical image analysis. Owing to their capacity to effectively learn global context and long-range dependencies from large volumes of image data, these models have emerged as a promising alternative to conventional Convolutional Neural Networks (CNNs), particularly in disciplines such as radiology and pathology. Specific to breast cancer, ViT-based approaches demonstrate high accuracy rates in the detection, classification, and segmentation of suspicious lesions from mammograms, histopathological sections, and ultrasound images.

The integration of artificial intelligence with established medical imaging techniques has pioneered the development of advanced early detection systems for a variety of health conditions [14–17]. The application of these systems spans a wide spectrum, from the diagnosis of retinal diseases [18, 19] to the detection of oncological pathologies such as breast [20–27], cervical [28–31], and brain cancers [32–35]. Drawing inspiration from these successful applications, the present study undertakes a comprehensive comparison of four prominent Transformer architectures (Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base) for the classification of breast ultrasound images. This analysis is focused on a detailed evaluation of the performance, diagnostic accuracy, and efficacy of the aforementioned models.

1.1. Related Works

The integration of artificial intelligence, particularly deep learning, has initiated a transformative era in breast cancer diagnostics, representing a significant leap forward in computational pathology. As noted by Katayama et al. [36], these technologies are increasingly automating crucial tasks for pathologists, such as tumor identification and classification, thereby enhancing the efficiency and throughput of pathology services. While Convolutional Neural Networks (CNNs) have historically been the foundational architecture in this domain, a paradigm shift is underway towards Vision Transformers (ViTs). The growing interest in ViTs stems from their architectural superiority in capturing long-range dependencies and global contextual information within whole-slide images (WSIs), a limitation of the localized receptive fields inherent in CNNs. This sentiment is echoed by multiple researchers, including Abimouloud et al. [37] and Balaha et al. [38], who highlight the potential of ViT-based systems to offer a more holistic image analysis, thus serving as powerful assistive tools that augment the diagnostic capabilities of medical professionals in modern breast cancer management.

The empirical evidence supporting the efficacy of ViT-based architectures is compelling and continues to grow. In a direct comparative study, Jahan et al. [39] demonstrated that a ViT-based model surpassed established CNNs like DenseNet-201 and MobileNetV2, achieving a superior accuracy of 96.74% for cancerous patch detection and 89.78% for subtype classification in WSIs. This highlights the model's robust feature extraction capabilities. The versatility of these architectures is further demonstrated by Boudouh and Bouakkaz, [40] who engineered a novel hybrid model combining a ViT++ branch with a VGG16 CNN branch. Their approach yielded an exceptional accuracy of 99.22% for the classification of breast calcifications in mammograms, showcasing the synergistic potential of fusing transformer and convolutional features. Further reinforcing these findings, Balaha et al. [38] developed a CAD framework utilizing ViTs that achieved state-of-the-art performance exceeding 97% accuracy. Notably, their work also integrated interpretability methods like SHAP, addressing the "black box" issue and enhancing the clinical trustworthiness of the model's predictions.

Despite their impressive performance, the practical deployment of standard ViT models is often hindered by significant challenges, including a substantial data dependency, high parametric complexity, and intensive computational resource requirements for training. Abimouloud et al. [37, 41] extensively discuss these limitations, which are particularly pronounced in medical imaging where datasets can be limited and

complex. In response, a prominent research trend has been the development of lightweight and hybrid ViT-CNN architectures designed to mitigate these issues. In one study, Abimouloud et al. proposed and validated several low-weight systems (ViT, CCT, MVIT), which achieved high accuracy (up to 98.64%) while minimizing computational overhead. In a subsequent investigation, the same research group introduced the TokenMixer, a sophisticated hybrid architecture inspired by ConvMixer and TokenLearner models. This model optimizes the feature extraction pipeline by strategically tokenizing input patches, resulting in reduced training times and fewer parameters while attaining a remarkable 97.02% accuracy in binary classification. Such innovations are critical for advancing the clinical viability of ViT-based systems, making them more efficient, accessible, and practical for real-world diagnostic applications.

2. Methods and Methodology

2.1. Dataset

The empirical basis of this study is the "Breast Ultrasound Images Dataset," made publicly available by sabahesaraki via the Kaggle platform [42]. Comprising a total of 780 ultrasound (US) images of pathologically confirmed benign, malignant, and normal breast tissue, this dataset provides a robust foundation for evaluating the capability of the developed models to discriminate between distinct tissue patterns. To ensure standardized and reproducible model development processes and to reliably assess generalization capability while minimizing the risk of overfitting, the dataset was partitioned using a stratified methodology. Accordingly, the data was allocated into training (70%, n=545), validation (10%, n=70), and independent test sets (20%, n=158), with the latter reserved for the final performance evaluation of the models. This stratification ensured that the class distribution within each subset mirrored the proportions of the original dataset, resulting in a training set comprising 305 benign, 147 malignant, and 93 normal images; a validation set with 43 benign, 21 malignant, and 13 normal images; and a test set containing 89 benign, 42 malignant, and 27 normal images. The specifics of this data partitioning are detailed in Table 1.

Table 1: Detailed categorical distribution of the "Breast Ultrasound Images Dataset" across training, validation, and testing subsets using a 70%–10%–20% stratified split ratio.

Classes	**Train**	**Validation**	**Test**	**Total**
Bening	305	43	89	437
Malignant	147	21	42	210
Normal	93	13	27	133
Total	**545**	**77**	**158**	**780**

To visually elucidate the inter-class morphological differences and imaging characteristics within the dataset, representative ultrasound examples from each category (benign, malignant, and normal) are presented in Figure 1. These images illustrate the structural dichotomy between the regular margins and homogeneous internal structure typically associated with benign lesions, and the irregular borders, spiculated extensions, and heterogeneous internal echoes characteristic of malignant lesions. Concurrently, the typical fibroglandular and adipose tissue patterns of normal breast parenchyma are also discernible. The presented exemplars not only highlight the discriminative pathological characteristics but also embody the technical challenges inherent to ultrasound imaging that the models must overcome, such as speckle noise and low contrast. Consequently, this visual presentation illuminates the fundamental features that the models are required to learn, while concurrently providing a comprehensive insight into the diversity and complexity inherent within the dataset.

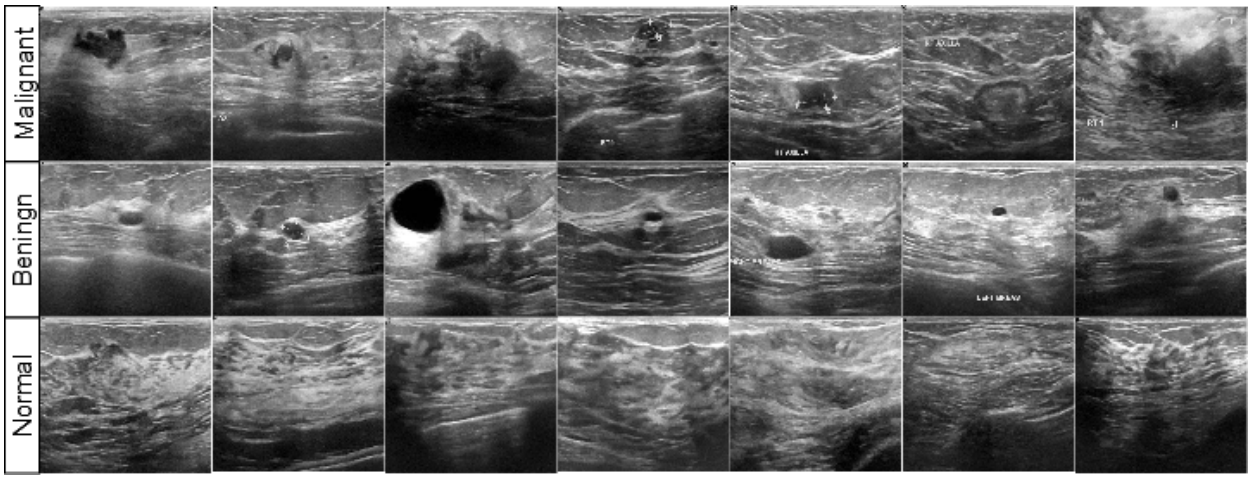

Figure 1: Representative ultrasonography samples showcasing the distinctive morphological features of Normal, Benign (regular margins), and Malignant (irregular/spiculated borders) breast tissue patterns.

2.2. Data Augmentation

In this study, to mitigate the risk of overfitting a prevalent challenge with limited medical datasets and to enhance generalization performance, a dynamic data augmentation strategy was incorporated into the training pipeline [43, 44]. Given the nature of the classification task, mask files (mask.png), initially provided for segmentation purposes, were excluded from the scope of the analysis. During training, a series of random transformations were applied to each image on-the-fly. Initially, each image was randomly cropped to a scale of 8% to 100% of its original area (scale: [0.08, 1.0]) with an aspect ratio maintained between 0.75 and 1.33 (ratio: [0.75, 1.33]). Subsequently, it was resized to a standard resolution of 224x224 pixels (img-size: 224) using a randomly selected interpolation method (train-interpolation: random). These geometric augmentations were supplemented by horizontal flipping, applied with a 50% probability (hflip: 0.5), whereas vertical flipping was not employed (vflip: 0.0). The photometric augmentations involved random adjustments to brightness, contrast, saturation, and hue, with a jitter factor of 0.4 (color-jitter: 0.4). This dynamic and multi-faceted augmentation strategy exposes the model to a different variation of the data during each training epoch, thereby inhibiting memorization and reinforcing the model's capacity for robust and reliable inference on previously unseen data.

2.3. Model Architecture

A recent paradigm shift in computer vision has been catalyzed by the adaptation of the revolutionary Transformer architecture, originally conceived for natural language processing. The pioneering model in this transition, the Vision Transformer (ViT) [45], re-envisioned image analysis by deconstructing an image into a sequence of flattened patches, treating them as tokens. This approach eschews the inductive biases of convolutional layers, instead leveraging a global self-attention mechanism to capture long-range dependencies across the entire image. While this enables a powerful holistic understanding, particularly when trained on massive datasets, the global attention mechanism incurs a significant computational burden and struggles with data efficiency. Addressing this latter issue, the Data-efficient Image Transformer (DeiT) [46] was introduced, employing a knowledge distillation strategy where a "student" model learns from a pre-trained "teacher" network. This method allows DeiT to achieve or even surpass ViT's performance without the need for vast training corpora. The DeiT architecture, shown in Figure 2, improves data efficiency by incorporating a 'distillation token' alongside the standard class token and implementing a teacher-student learning strategy.

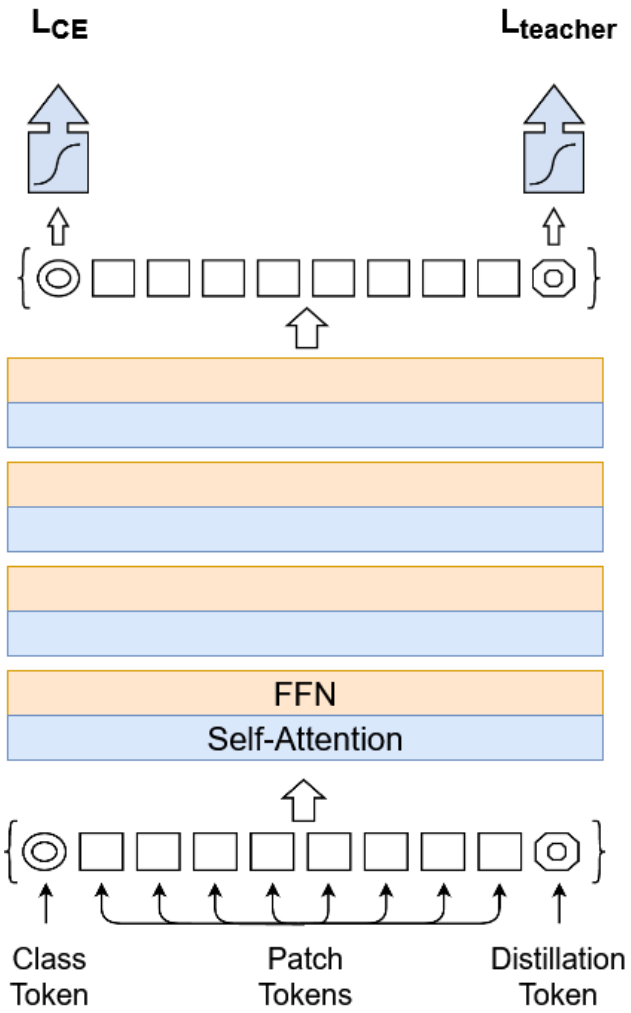

Figure 2: Detailed schematic of the Data-efficient Image Transformer (DeiT) architecture, highlighting the integration of the distillation token and the teacher-student knowledge transfer mechanism.

To tackle ViT's inherent scalability and computational challenges, the Swin Transformer [47] proposed a more pragmatic, hierarchical architecture. It introduces locality by confining the self-attention mechanism to non-overlapping windows, drastically reducing computational cost. To restore a global receptive field, a subsequent window shifting mechanism facilitates cross-window connections in deeper layers, creating a hierarchical feature representation analogous to that of CNNs. This design has proven superior for complex downstream tasks like object detection and semantic segmentation. Concurrently, a different approach to pre-training was inspired by the success of BERT in NLP, leading to BEiT (Bidirectional Encoder representations from Image Transformers) [48]. BEiT is pre-trained using a self-supervised task known as masked image modeling, where it learns to reconstruct original image patches from a corrupted version with masked regions. This process compels the model to learn robust, high-level semantic representations of image structure before it is fine-tuned for a specific task. Collectively, these pioneering architectures represent a significant evolution beyond traditional convolutional approaches, charting a new course for state-of-the-art image analysis.

3. Results and Discussion

In this study, the performance of four distinct Vision Transformer-based models (Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base) was comparatively analyzed for the classification of breast ultrasound images, considering various metrics and computational costs. The obtained results are summarized in Table 2. Following the evaluations, the DeiT-Base model was observed to exhibit a distinct superiority over all other models, achieving the highest performance. DeiT-Base attained impressive values, including an accuracy of 94.30%, a precision of 94.05%, a recall of 93.65%, and an F1-score of 93.85%. In contrast, the ViT-Base and Swin-Base models yielded competitive yet inferior results, with accuracy rates of 89.87% and 88.61%, respectively. The BEiT-Base model, with an accuracy of 66.46%, exhibited a performance substantially below expectations.

Table 2: Comprehensive performance evaluation and computational complexity analysis of Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base models based on diagnostic metrics and GFLOPs.

Models	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Params**	**GFLOPs**
Swin-Base	88.61	88.38	87.72	87.97	86.75	30.3375
ViT-Base	89.87	89.56	88.05	88.77	85.80	33.7257
Deit-Base	94.30	94.05	93.65	93.85	85.80	33.7257
Beit-Base	66.46	47.97	47.71	45.43	85.76	25.3294

Upon examining the performance disparities among the models, the success of DeiT-Base is attributed to its knowledge distillation strategy, which trains the ViT architecture in a data-efficient manner. This ”teacher-student” approach enables the model to learn richer and more generalizable features from a limited dataset, offering a significant advantage over standard training methods, particularly in data-scarce domains like medical imaging. While the performances of ViT-Base and Swin-Base were quite close, a notable trade-off between computational cost and performance was observed. ViT-Base offered slightly higher accuracy (1%), whereas the Swin-Base, owing to its hierarchical and windowed attention mechanism, was found to operate with a lower computational load (30.34 GFLOPs vs. 33.73 GFLOPs). This indicates that the Swin architecture could be a more pragmatic option in resource-constrained settings. The confusion matrices derived from the architectures of the employed models are presented in Figure 3.

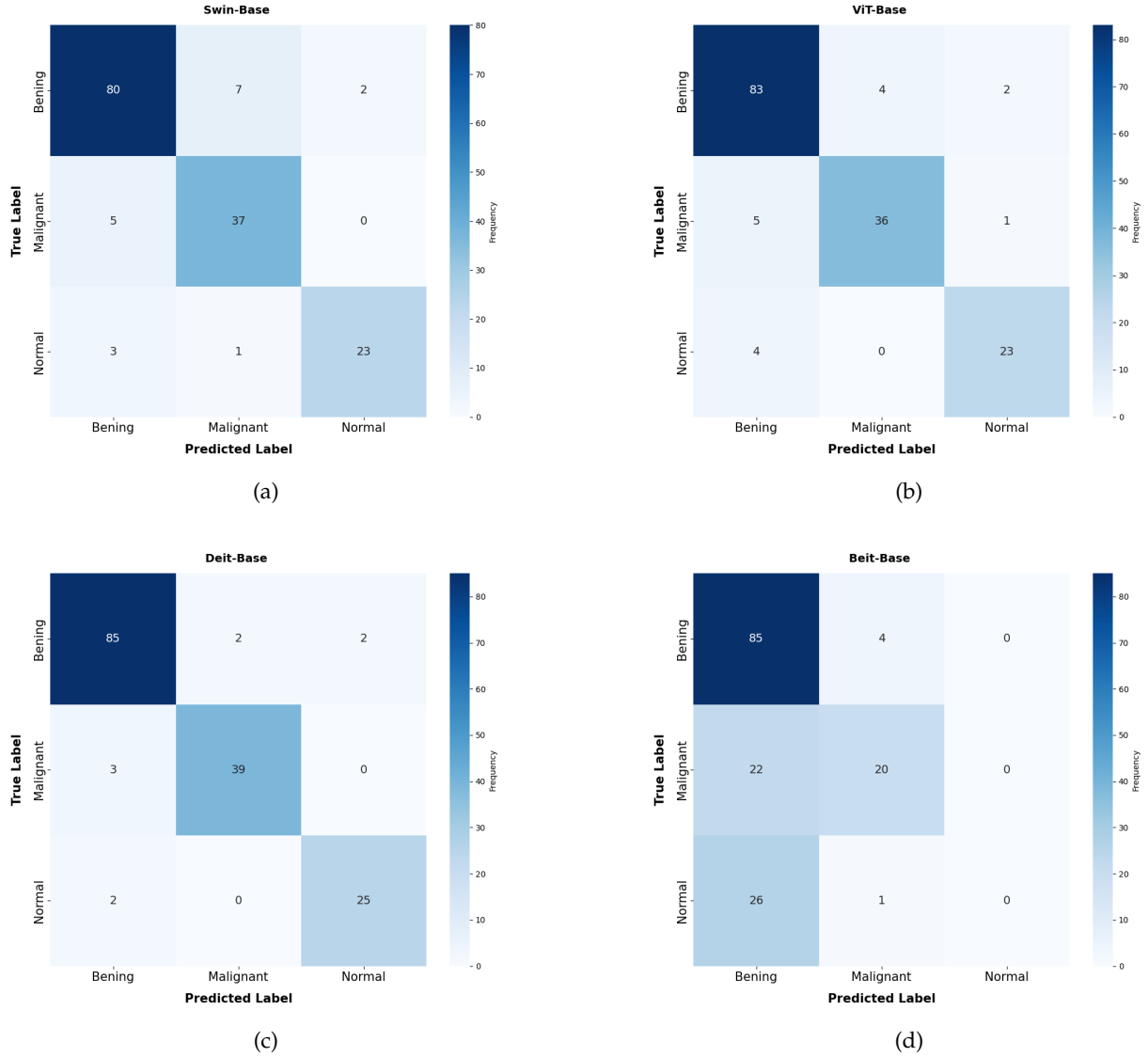

Figure 3: Visualization of multi-class classification performance via confusion matrices for the four evaluated Transformer architectures on the independent test dataset.

One of the most striking findings was the suboptimal performance exhibited by the BEiT-Base model. The model's remarkably low F1-score of 45.43% suggests an ineffective transfer of the features learned through self-supervised pre-training to this specific medical image classification task. The hypothesis can be advanced that although BEiT learns powerful semantic representations via "masked image modeling" on natural images, these representations failed to adapt to the unique characteristics of ultrasound images, such as their distinct texture patterns, speckle noise, and low contrast. Despite being the most efficient model with the lowest GFLOPs, this weak performance underscores that computational efficiency alone is insufficient for achieving high diagnostic accuracy.

This investigation demonstrates that for the classification of breast ultrasound images, the training strategy plays a role as critical as the model architecture. The superior success achieved by DeiT-Base through its knowledge distillation method highlights the potential of such data-efficient training approaches for future medical image analysis studies. Future investigations could explore whether the performance

of self-supervised models like BEiT can be enhanced through pre-training on large, in-domain medical datasets.

4. Conclusions

This study aimed to comprehensively analyze the performance of four distinct Vision Transformer-based architectures (Swin-Base, ViT-Base, DeiT-Base, and BEiT-Base) for the automatic classification of breast ultrasound images. The findings unequivocally demonstrate that the DeiT-Base model, with an accuracy of 94.30%, significantly outperformed all other tested approaches. It is reasonable to attribute this superiority to the effectiveness of the knowledge distillation strategy employed by DeiT in learning more robust and generalizable features, particularly when working with limited medical datasets. Although the ViT-Base and Swin-Base models exhibited acceptable performance, they fell short of DeiT's success. Conversely, the weak performance of the BEiT-Base model illuminates the challenges of directly transferring representations learned via self-supervision on general-purpose natural images to medical imaging modalities, like ultrasound, which possess unique noise and texture characteristics. This situation serves as an indicator of how critical training strategies, tailored to the nature of the data, are, in addition to the choice of model architecture.

The execution of this study on a single public dataset should be acknowledged as a limitation, and the findings warrant validation with more extensive datasets collected from diverse clinical environments. Several key directions are proposed for future research. First, pre-training models such as BEiT using large, in-domain medical image archives instead of natural images holds substantial potential for significantly enhancing their performance. Second, inspired by the success of DeiT, integrating knowledge distillation techniques into more computationally efficient architectures like the Swin Transformer may yield an optimal balance between accuracy and efficiency. Ultimately, the integration of such high-performance models into clinical decision support systems, in tandem with explainability methods, is poised to maximize the practical value of these technologies in early diagnostic processes and promote their adoption by clinicians.

5. Declaration of AI Use

We declare that large language models (LLMs) were used solely to assist in improving the language quality, clarity, and readability of this manuscript. The AI tools were not employed in the generation of scientific ideas, experimental design, data analysis, result interpretation, or conclusion formulation. All scientific content and final decisions were made by the authors, who take full responsibility for the accuracy, originality, and integrity of the work.

References

[1] Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., Mutebi, M., Garvey, G., Soerjomataram, I., and Fidler-Benaoudia, M. M. (2025). Global Patterns and Trends in Breast Cancer Incidence and Mortality across 185 Countries. Nature Medicine, 31(4), 1154–1162.

[2] Xiong, X., Zheng, L. W., Ding, Y., Chen, Y. F., Cai, Y. W., Wang, L. P., Huang, L., Liu, C. C., Shao, Z. M., and Yu, K. D. (2025). Breast Cancer: Pathogenesis and Treatments. Signal Transduction and Targeted Therapy, 10(1), 1–33.

[3] Obeagu, E. I., and Obeagu, G. U. (2024). Breast Cancer: A Review of Risk Factors and Diagnosis. Medicine (United States), 103(3), e36905.

[4] Kiani, P., Vatankhahan, H., Zare-Hoseinabadi, A., Ferdosi, F., Ehtiati, S., Heidari, P., Dorostgou, Z., Movahedpour, A., Baktash, A., Rajabivahid, M., and Khatami, S. H. (2025). Electrochemical Biosensors for Early Detection of Breast Cancer. Clinica Chimica Acta, 564, 119923.

[5] Alshawwa, I. A., El-Mashharawi, H. Q., Salman, F. M., Al-Qumboz, M. N. A., Abunasser, B. S., and Abu-Naser, S. S. (2024). Advancements in Early Detection of Breast Cancer: Innovations and Future Directions. PhilPapers.

[6] Begum, M. M. M. M., Gupta, R., Sunny, B., and Lutfor, Z. L. (2024). Advancements in Early Detection and Targeted Therapies for Breast Cancer: A Comprehensive Analysis. Asia Pacific Journal of Cancer Research, 1(1), 4–13.

[7] Katsika, L., Boureka, E., Kalogiannidis, I., Tsakiridis, I., Tirodimos, I., Lallas, K., Tsimtsiou, Z., and Dagklis, T. (2024). Screening for Breast Cancer: A Comparative Review of Guidelines. Life, 14(6), 777.

[8] Trentham-Dietz, A., Chapman, C. H. H., Jayasekera, J., Lowry, K. P., Heckman-Stoddard, B. M., Hampton, J. M., Caswell-Jin, J. L., Gangnon, R. E., Lu, Y., Huang, H., Stein, S., Sun, L., Gil Quessep, E. J., Yang, Y., Lu, Y., Song, J., Muñoz, D. F., Li, Y., Kurian, A. W., Kerlikowske, K., O'Meara, E. S., Sprague, B. L., Tosteson, A. N. A., Feuer, E. J., Berry, D., Plevritis, S. K., Huang, X., De Koning, H. J., Van Ravesteyn, N. T., Lee, S. J., Alagoz, O., Schechter, C. B., Stout, N. K., Miglioretti, D. L., and Mandelblatt, J. S. (2024). Collaborative Modeling to Compare Different Breast Cancer Screening Strategies: A Decision Analysis for the US Preventive Services Task Force. JAMA, 331(22), 1947–1960.

[9] Abeelh, E. A., and AbuAbeileh, Z. (2024). Comparative Effectiveness of Mammography, Ultrasound, and MRI in the Detection of Breast Carcinoma in Dense Breast Tissue: A Systematic Review. Cureus, 16(4).

[10] Iacob, R., Iacob, E. R., Stoicescu, E. R., Ghenciu, D. M., Cocolea, D. M., Constantinescu, A., Ghenciu, L. A., and Manolescu, D. L. (2024). Evaluating the Role of Breast Ultrasound in Early Detection of Breast Cancer in Low- and Middle-Income Countries: A Comprehensive Narrative Review. Bioengineering, 11(3), 262.

[11] Gordon, P. B., Warren, L. J., and Seely, J. M. (2025). Cancers Detected on Supplemental Breast Ultrasound in Women with Dense Breasts: Update from a Canadian Centre. Canadian Association of Radiologists Journal, 08465371251318578.

[12] Vogel-Minea, C. M., Bader, W., Blohmer, J.-U., Duda, V., Eichler, C., Fallenberg, E., Farrokh, A., Golatta, M., Gruber, I., Hackelöer, B.-J., and Others (2025). Best Practice Guidelines–DEGUM Recommendations on Breast Ultrasound. Ultraschall in der Medizin–European Journal of Ultrasound, 46(03), 245–258.

[13] Rana, A. S., Rafique, J., and Riffat, H. (2024). Advances in Breast Ultrasound Imaging: Enhancing Diagnostic Precision and Clinical Utility. In Latest Research on Breast Cancer–Molecular Insights, Diagnostic Advances and Therapeutic Innovations. IntechOpen.

[14] Paçal, İ. (2025). Diagnostic Analysis of Various Cancer Types with Artificial Intelligence. Duvar Design.

[15] Cakmak, Y., and Maman, A. (2025). Deep Learning for Early Diagnosis of Lung Cancer. Computational Systems and Artificial Intelligence, 1(1), 20–25.

[16] Cakmak, Y. (2025). Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders. Computational Systems and Artificial Intelligence, 1(1), 8–14.

[17] Cakmak, Y., and Paçal, I. (2025). AI-Driven Classification of Anemia and Blood Disorders Using Machine Learning Models. Computers and Electronics in Medicine, 2(2), 43–52.

[18] Kayadibi, I., and Güraksın, G. E. (2023). An Early Retinal Disease Diagnosis System Using OCT Images via CNN-Based Stacking Ensemble Learning. International Journal for Multiscale Computational Engineering.

[19] Kayadibi, İ., Güraksın, G. E., and Köse, U. (2023). A Hybrid R-FTCNN Based on Principal Component Analysis for Retinal Disease Detection from OCT Images. Expert Systems with Applications.

[20] Coşkun, D., Karaboğa, D., Baştürk, A., Akay, B., Nalbantoğlu, Ö. U., Doğan, S., Paçal, İ., and Karagöz, M. A. (2023). A Comparative Study of YOLO Models and a Transformer-Based YOLOv5 Model for Mass Detection in Mammograms. Turkish Journal of Electrical Engineering and Computer Sciences, 31(7), 1294–1313.

[21] Paçal, I., and Attallah, O. (2025). InceptionNeXt-Transformer: A Novel Multi-Scale Deep Feature Learning Architecture for Multimodal Breast Cancer Diagnosis. Biomedical Signal Processing and Control, 110, 108116.

[22] Paçal, İ. (2022). Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images. Journal of the Institute of Science and Technology, 12(4), 1917–1927.

[23] Işık, G., and Paçal, İ. (2024). Few-Shot Classification of Ultrasound Breast Cancer Images Using Meta-Learning Algorithms. Neural Computing and Applications, 36(20), 12047–12059.

[24] Cakmak, Y., and Paçal, I. (2025). Enhancing Breast Cancer Diagnosis: A Comparative Evaluation of Machine Learning Algorithms Using the Wisconsin Dataset. Journal of Operations Intelligence, 3(1), 175–196.

[25] Cakmak, Y., Safak, S., Bayram, M. A., and Paçal, I. (2024). Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection. Computer and Decision Making: An International Journal, 1, 84–102.

[26] Cakmak, Y., and Zeynalov, J. (2025). A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms. Artificial Intelligence in Applied Sciences, 1(1), 28–34.

[27] Cakmak, Y., and Pacal, N. (2025). Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures. Artificial Intelligence in Applied Sciences, 1(1), 13–19.

[28] Karaman, A., Paçal, I., Baştürk, A., Akay, B., Nalbantoğlu, U., Coşkun, S., Şahin, Ö., and Karaboğa, D. (2023). Robust Real-Time Polyp Detection System Design Based on YOLO Algorithms by Optimizing Activation Functions and Hyper-Parameters with Artificial Bee Colony (ABC). Expert Systems with Applications, 221, 119741.

[29] Karaman, A., Karaboğa, D., Paçal, I., Akay, B., Baştürk, A., Nalbantoğlu, U., Coşkun, S., and Şahin, Ö. (2023). Hyper-Parameter Optimization of Deep Learning Architectures Using Artificial Bee Colony (ABC) Algorithm for High Performance Real-Time Automatic Colorectal Cancer (CRC) Polyp Detection. Applied Intelligence, 53(12), 15603–15620.

[30] Paçal, İ. (2024). MaxCerVixT: A Novel Lightweight Vision Transformer-Based Approach for Precise Cervical Cancer Detection. Knowledge-Based Systems, 289, 111482.

[31] Paçal, I., and Kılıcarslan, S. (2023). Deep Learning-Based Approaches for Robust Classification of Cervical Cancer. Neural Computing and Applications, 35(25), 18813–18828.

[32] Ince, S., Kunduracioglu, I., Algarni, A., Bayram, B., and Paçal, I. (2025). Deep Learning for Cerebral Vascular Occlusion Segmentation: A Novel ConvNeXtV2 and GRN-Integrated U-Net Framework for Diffusion-Weighted Imaging. Neuroscience, 574, 42–53.

[33] Paçal, I. (2024). A Novel Swin Transformer Approach Utilizing Residual Multi-Layer Perceptron for Diagnosing Brain Tumors in MRI Images. International Journal of Machine Learning and Cybernetics, 15(9), 3579–3597.

[34] Paçal, I., Akhan, O., Tuna Deveci, R., and Deveci, M. (2025). NeXtBrain: Combining Local and Global Feature Learning for Brain Tumor Classification. Brain Research, 1863, 149762.

[35] Paçal, I., Celik, O., Bayram, B., and Cunha, A. (2024). Enhancing EfficientNetv2 with Global and Efficient Channel Attention Mechanisms for Accurate MRI-Based Brain Tumor Classification. Cluster Computing, 27(8), 11187–11212.

[36] Katayama, A., Aoki, Y., Watanabe, Y., Horiguchi, J., Rakha, E. A., and Oyama, T. (2024). Current Status and Prospects of Artificial Intelligence in Breast Cancer Pathology: Convolutional Neural Networks to Prospective Vision Transformers. International Journal of Clinical Oncology, 29(11), 1648–1668.

[37] Abimouloud, M. L., Bensid, K., Elleuch, M., Ben Ammar, M., and Kherallah, M. (2024). Vision Transformer Based Convolutional Neural Network for Breast Cancer Histopathological Images Classification. Multimedia Tools and Applications, 83(39), 86833–86868.

[38] Balaha, H. M., Ali, K. M., Gondim, D., Ghazal, M., and El-Baz, A. (2025). Harnessing Vision Transformers for Precise and Explainable Breast Cancer Diagnosis. Lecture Notes in Computer Science, 15311, 191–206.

[39] Jahan, I., Chowdhury, M. E. H., Vranic, S., Al Saady, R. M., Kabir, S., Pranto, Z. H., Mim, S. J., Nobi, S. F., and Nobi, S. F. (2025). Deep Learning and Vision Transformers-Based Framework for Breast Cancer and Subtype Identification. Neural Computing and Applications, 37(16), 9311–9330.

[40] Boudouh, S. S., and Bouakkaz, M. (2024). Advancing Precision in Breast Cancer Detection: A Fusion of Vision Transformers and CNNs for Calcification Mammography Classification. Applied Intelligence, 54(17–18), 8170–8183.

[41] Abimouloud, M. L., Bensid, K., Elleuch, M., Ben Ammar, M., and Kherallah, M. (2025). Advancing Breast Cancer Diagnosis: Token Vision Transformers for Faster and Accurate Classification of Histopathology Images. Visual Computing for Industry, Biomedicine, and Art, 8(1), 1–27.

[42] Breast Ultrasound Images Dataset (BUSI). Available at: https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset.

[43] Wang, Z., Wang, P., Liu, K., Wang, P., Fu, Y., Lu, C. T., Aggarwal, C. C., Pei, J., and Zhou, Y. (2025). A Comprehensive Survey on Data Augmentation. IEEE Transactions on Knowledge and Data Engineering.

[44] Mumuni, A., Mumuni, F., and Gerrar, N. K. (2024). A Survey of Synthetic Data Augmentation Methods in Machine Vision. Machine Intelligence Research, 21(5), 831–869.

[45] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021).

[46] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jégou, H. (2020). Training Data-Efficient Image Transformers & Distillation through Attention. Proceedings of Machine Learning Research, 139, 10347–10357.

[47] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. Proceedings of the IEEE International Conference on Computer Vision, 9992–10002.

[48] Bao, H., Dong, L., Piao, S., and Wei, F. (2021). BEiT: BERT Pre-Training of Image Transformers. In Proceedings of the 10th International Conference on Learning Representations (ICLR 2022).