2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)
29-30 November 2025, Eastern University, Dhaka, Bangladesh

2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) | 979-8-3315-6105-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/BECITHCON69222.2025.11504068

Lightweight Hybrid Transformer System for Robust and Explainable Multi-Modality Breast Cancer Recognition

Mahfuzur Rahman[1], Rasel Mahmud Jewel[2], Md Imranul Hoque Bhuiyan[3], Sanjida Akter[4], Istiak Kabir[5], Abdullah Al Sakib[6], Shafiur Rahman[7], Mostafizur Rahman Shakil[8]
[1]*School of Computing and Digital Media, London Metropolitan University, London N7 8DB, UK*
[2]*College of Business, Westcliff University, CA 92614, USA*
[3]*Department of Business Analytics, International American University, CA 90010, USA*
[4]*Department of Mathematics and Natural Sciences, BRAC University, Dhaka 1212, Bangladesh*
[5]*Department of Computer Science, Westcliff University, CA 92614, USA*
[6]*Department of Information Technology, Westcliff University, CA 92614, USA*
[7]*Dept. of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh*
[8]*Department of Engineering Management, Westcliff University, CA 92614, USA*
Email: [1]mahfuzurrahman63036@gmail.com, [2]r.jewel.198@westcliff.edu, [3]imranrafi1111@gmail.com, [4]sanjidairen261@gmail.com, [5]istiak2221@gmail.com, [6]sakibabdulla685@gmail.com, [7]shafiur016@gmail.com, [8]mostafizurrahmanshakil@gmail.com

***Abstract*— Breast cancer is a major global health concern, making early detection crucial for improving survival rates. Deep learning methods in this field face challenges such as class imbalance, varying imaging types, and clinical interpretability. This study introduces a lightweight transformer-based model, EFormer-EA, aimed at breast cancer classification across different imaging modalities. Our hybrid architecture uses EfficientFormerV2 for local feature extraction and External Attention for modeling global dependencies. We applied it to two public datasets: BreakHis, containing 7,909 histopathology images at four magnifications, and BUSI, with 830 ultrasound images across three classes. Preprocessing included modality-specific normalization, histogram equalization, and GPU-accelerated augmentation, with a class-weighted loss to address class imbalance. The EFormer-EA model achieved impressive results: an F1 score of 98.27% and a Matthew's correlation coefficient of 96.15 on BreakHis, and an F1 score of 98.46% and a PR-AUC of 99.43 on BUSI, outperforming existing models. We also incorporated Grad-CAM into a web application for real-time, explainable diagnosis. However, the study has limitations, including sensitivity to external memory sizes and dependence on just two datasets. Future work will focus on multi-center validation, edge deployment, and integration with federated learning.**

Keywords—Breast cancer, vision transformer, explainable AI, medical imaging, diagnostic tools

I. INTRODUCTION

Breast cancer is one of the leading causes of cancer-related deaths among women worldwide. Early detection is crucial for improving patient survival rates and treatment outcomes. According to the World Health Organization, breast cancer accounts for approximately 2.3 million new cases and 685,000 deaths each year [1], underscoring the urgent need for accurate, efficient, and accessible diagnostic solutions. Traditional diagnostic methods, such as histopathology and ultrasound, are considered the gold standard. However, their effectiveness is often limited by manual interpretation, variability among observers, and significant time demands on clinicians [2]. These challenges highlight the need for automated, accurate, and interpretable diagnostic systems to support clinical decision-making.

In recent years, deep learning (DL) has shown great potential in enhancing breast cancer diagnosis through automated feature extraction and classification from medical images. Convolutional neural networks (CNNs) have been widely used because of their strong ability to learn local features, while Vision Transformers (ViTs) have emerged as powerful alternatives that excel at capturing long-range spatial dependencies [3]. However, several challenges remain, including data imbalance, overfitting, domain shift, high computational requirements, and limited interpretability [4], issues that impede real-world clinical deployment. Additionally, most existing studies focus on a single imaging modality, which restricts their applicability across diverse diagnostic scenarios [5].

Various approaches have been proposed to overcome these limitations. Hybrid CNN-based models have achieved high accuracy but often come with significant computational overhead and limited scalability. Generative and optimization-based techniques have improved performance with imbalanced datasets, but they tend to be computationally intensive and do not adequately address issues of interpretability [6]. While ViT-based models show promise in capturing global context, they struggle with overfitting on small datasets and lack robust explainable AI (XAI) integration. Furthermore, most existing systems are not designed for real-time web-based applications, which are essential for improving accessibility and clinical adoption [7].

This study identifies a critical gap in research: the need for lightweight, accurate transformer-based architecture that can handle multiple imaging modalities, address class imbalance, and offer interpretable, real-time clinical decision

support. The integration of XAI into diagnostic platforms remains limited, affecting transparency and trust in clinical applications. The main objectives are to: (i) develop a DL system for reliable breast cancer diagnoses across various imaging modalities; (ii) address class imbalance and generalization issues in medical image classification; (iii) enhance interpretability of AI diagnostic tools; (iv) evaluate model performance using metrics suitable for imbalanced and multi-class datasets.

To achieve these goals, we propose EFormer-EA, a hybrid transformer architecture that utilizes EfficientFormerV2 for local feature extraction and External Attention for global dependencies. We preprocessed two public datasets with modality-specific normalization, histogram equalization, and GPU-accelerated data augmentation to improve robustness and manage class imbalance. A class-weighted loss function addresses misclassifications of minority classes. The model is trained and evaluated with multiple metrics for balanced performance. Following are our contributions:

- Introduced EFormer-EA is a hybrid transformer-based model that combines EfficientFormerV2 stages with External Attention, aiming to achieve a balance between computational efficiency and high diagnostic accuracy.
- Created a robust training pipeline with specific preprocessing for each modality, accelerated GPU augmentation, and class-weighted loss functions to enhance performance despite data imbalance.
- Integrated Grad-CAM into a potential user-friendly web application for transparent and interpretable diagnosis.
- Achieved state-of-the-art results on the BreakHis and BUSI datasets, surpassing previous models and demonstrating strong adaptability across imaging modalities.

This paper is organized as follows: Section II reviews related work on convolutional, transformer-based, and generative/optimization approaches for breast cancer diagnosis. Section III explains the methodology, covering dataset preprocessing, model architecture, and training configuration. Section IV presents experimental results and performance analysis. Section V discusses implications, limitations, and future research directions. Finally, Section VI concludes the paper.

II. Related Works

A growing body of work illustrates the broad application of artificial intelligence, spanning legal informatics, cybersecurity solutions, positioning technologies, and bibliometric analysis [8], [9], [10], [11]. Parallel efforts have investigated its integration with decentralized frameworks such as blockchain consensus, atomic swap protocols, and metaverse asset management [12], [13], [14]. Complementary research highlights AI's contributions to agricultural automation, intelligent navigation, medical imaging, and machine reliability prediction [15], [16], [17], [18].

Historically, CNN-based models have dominated breast cancer image classification due to their strong feature extraction capabilities. Chakravarthy et al. [19] introduced a hybrid model that combines VGG16, ResNet50, and DenseNet121, achieving over 98% accuracy on the CBIS-DDSM and INbreast datasets. However, this model struggled with malignancy classification and required significant computational resources. Wani et al. [20] proposed BC2, a CNN-LightGBM (LGBM) pipeline with SHAP-based interpretability, achieving 98.3% accuracy on SEER data but showing reduced generalization when faced with domain shifts. Ajlouni et al. [21] employed a CNN-Self-Organizing Map (SOM) system for Tumor, Node, Metastasis (TNM) classification from breast MRI, achieving 98% staging accuracy. However, the lack of validation with external datasets limits its clinical relevance.

To tackle data scarcity and class imbalance, researchers have explored generative and metaheuristic approaches. Munteanu et al. [22] utilized GAN-based augmentation in a CNN-U-Net, which improved learning for minority classes on the BUSI dataset but plateaued at 86% accuracy. Jabeen et al. [23] developed a residual CNN enhanced by a quantum distribution optimizer and kernel canonical correlation analysis (CCA), achieving 96.5% accuracy on the INbreast dataset, albeit with high computational costs and extensive parameter tuning. Emam et al. [24] integrated a Lévy flight-based chaotic optimization algorithm into DenseNet121, reaching an impressive 99.87% accuracy on the DMR dataset, although this method remained computationally expensive and lacked interpretability and privacy guarantees, which reduces its clinical applicability. ViTs have shown strong capabilities in capturing global spatial dependencies in medical imaging. Srivastava et al. [25] compared MaxViT, CvT, and CrossViT for Invasive Ductal Carcinoma (IDC) histopathology, achieving 92.12% accuracy, though performance declined on heterogeneous datasets. Hayat et al. [26] combined EfficientNetV2 with ViT, achieving 99.83% accuracy for binary classification and 98.10% for multiclass classification on the BreaKHis dataset, but they reported issues with overfitting and optimization challenges. Shiri et al. [27] proposed SupCon-ViT, a supervised contrastive ViT model, attaining 88.61% balanced accuracy; however, it lacked evaluation across diverse datasets, raising concerns about scalability.

Despite these advancements, the reviewed studies failed to use effective interpretable methods to clearly show the reasoning behind their predictions and did not offer a web-based application for real-time interpretability. This lack of integrated XAI and an interactive platform hinders transparency, trust, and practical use of these models in clinical settings, where both explainability and easy interaction for clinicians are essential for adoption.

III. Methodology

A. Data Description

Figure 1 illustrates the overall methodology of the study for multi-modality breast cancer classification. The process begins with data acquisition from two publicly available breast cancer imaging datasets (Figure 2): BreakHis [28] (histopathology) and BUSI [29] (ultrasound). BreakHis contains 7,909 histopathological images from 82 patients across four magnifications (40×, 100×, 200×, and 400×), including 2,480 benign and 5,429 malignant samples, with the majority being Ductal Carcinoma. The BUSI dataset comprises 830 ultrasound images from 600 patients,

categorized into normal (133), benign (487), and malignant (210). Sample images from each class are shown in Figure 2. For both datasets, the data is split into training, validation, and test sets in an 80:5:15 ratio. Following acquisition, modality-specific preprocessing steps such as resizing, normalization, histogram equalization, augmentation, and class imbalance handling are applied. The preprocessed data is then fed into the hybrid EFormer-EA model, which integrates EfficientFormerV2 for local feature extraction and External Attention for capturing global dependencies. Additionally, baseline transformer models are evaluated for comparison.

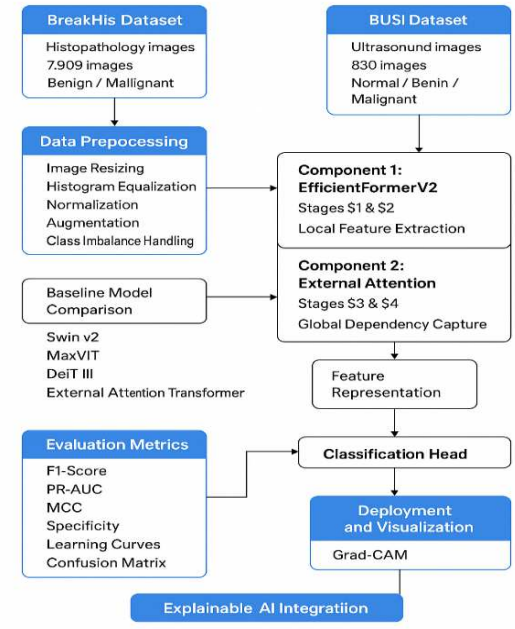

Fig. 1. Overview of the proposed methodology.

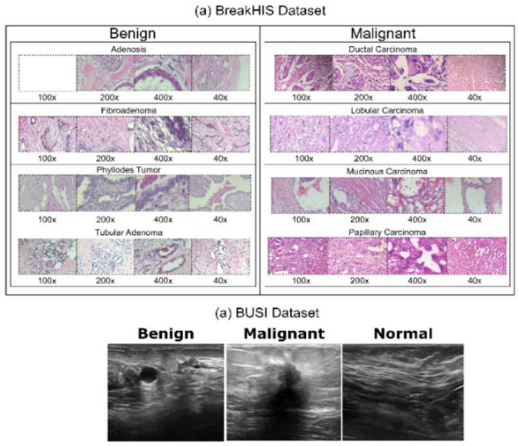

Fig. 2. Representative class sample from experimental datasets.

B. Data Preprocessing

All images were resized to 224 × 224 pixels for the experimental models. Histogram equalization enhanced intensity distribution, particularly in histopathological and ultrasound images. The BUSI dataset's grayscale images and segmentation masks were preserved for diagnostic detail. To reduce overfitting and improve robustness, data augmentation techniques were applied [30], including random rotations (±15°), flips, zooming (±10%), brightness and contrast adjustments, and elastic transformations. These augmentations were applied during training using Kornia's GPU-accelerated pipeline to avoid CPU bottlenecks and enhance training efficiency. To address class imbalance, a class-weighted loss function was used, assigning higher weights to minority classes to penalize misclassifications more heavily. All pixel intensities were normalized to [0, 1], and mean-variance normalization was performed using modality-specific statistics.

C. Base Models

We evaluated four advanced ViT variants as baseline models: Swin v2, MaxViT, DeiT III, and External Attention Transformer (EAT). Each was chosen for its compatibility with different imaging modalities. Swin v2 features a hierarchical design with shifted window-based attention, allowing for localized computation and the retention of long-range dependencies [31]. It excels in analyzing high-resolution images, where detecting microcalcifications and architectural distortions is critical. The hierarchical structure is defined as:

$$z^{l+1} = \mathrm{MLP}\left(\widehat{z^l} + \mathrm{SWA}\left(\widehat{z^l}\right)\right), \quad \widehat{z^l} = \mathrm{LN}\left(z^l\right)$$

Here, SWA denotes the shifted window attention and MLP is a position-wise feed-forward layer. Layer normalization (LN) is applied prior to attention and MLP blocks. Next, MaxViT combines MBConv blocks with both local window and grid-based attention, enabling fine-to-coarse modeling of spatial hierarchies [32], critical for histopathological image analysis at multiple magnifications (BreakHis dataset). Its core building block is defined as:

$$x' = \mathrm{GridAttn}\left(\mathrm{WinAttn}\left(\mathrm{MBConv}(x)\right)\right)$$

Here, MBConv captures local inductive bias through convolution, while WinAttn and GridAttn enhance long-range dependency modeling at multiple scales. Furthermore, DeiT III is a data-efficient transformer optimized for limited supervision via progressive token distillation and improved training recipes [33]. It is well suited for the low-sample, high-variance ultrasound dataset. The backbone applies token mixing through residual connections:

$$z^{l+1} = \mathrm{DropPath}\left(M\left(z^l\right)\right) + z^l$$

Here, $\mathrm{M}(\mathrm{z}^l)$ denotes a modified transformer block with enhanced regularization strategies. DeiT III performs robustly without relying on large-scale pretraining. Lastly, EAT introduces an external memory-based attention mechanism to reduce redundancy and improve computational efficiency. It is evaluated across both to assess multi-class robustness and modality transferability. The architecture integrates a learnable external memory ($\mathrm{M} \in \mathrm{R}^{\mathrm{d}\times\mathrm{k}}$) and is formulated as:

$$Y = X + \mathrm{ExternalAttn}(X, M)$$

Here, ($\mathrm{X} \in \mathrm{R}^{\mathrm{n}\times\mathrm{d}}$) are token embeddings. External memory enhances attention diversity while maintaining compact representations across imaging domains.

D. Proposed EFormer-EA

We propose EFormer-EA (Figure 3), a lightweight transformer architecture that combines EfficientFormerV2 stages with External Attention (EA). This model is specifically designed to retain local texture details that are important for histopathology while efficiently modeling global dependencies for ultrasound imaging.

The architecture consists of four stages. Stages S1 and S2 utilize EfficientFormerV2's MetaFormer blocks, which

include depthwise convolution (DWConv) token mixers and feed-forward networks (FFN). These stages perform aggressive token downsampling, reducing both the sequence length and computational load. In contrast, stages S3 and S4 replace the token mixer with EA, which integrates a learnable external memory ($M \in R^{k \times d}$) to diversify attention and reduce redundancy with linear complexity ($O(Nk)$), where (N) is the token count and ($k \ll N$). Formally, the EfficientFormerV2 block can be expressed as:

$$X' = X + \phi\left(\mathrm{DWConv}\left(\mathrm{LN}(X)\right)\right), Y = X' + \mathrm{FFN}\left(\mathrm{LN}(X')\right)$$

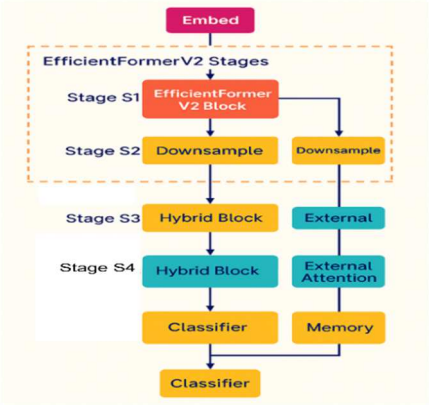

Fig. 3. Proposed Eformer-EA model.

while the EA-enhanced block is represented as:

$$Y = X + \mathrm{ExternalAttn}(X, M)$$

Here, the ExternalAttn function calculates attention weights using external keys and values, improving cross-token interactions without the quadratic overhead associated with self-attention.

The EFormer-EA combines EfficientFormerV2's early token downsampling and depthwise convolution for efficient local feature extraction with late-stage External Attention for O(Nk) global dependency modeling. This hybrid design enhances accuracy and efficiency in histopathology and ultrasound while decreasing FLOPs, latency, and memory usage compared to other models [34]. It includes modality-specific normalization, class-weighted loss, and GPU-accelerated augmentation to improve generalization under class imbalance. However, it is sensitive to external memory size k, requires tuning at high resolutions, and is more expensive than lightweight CNNs on ultra-low-power devices.

E. Evaluation

We evaluated performance using several metrics: F1-Score, PR-AUC, MCC, and Specificity. These metrics help address class imbalance and ensure reliable diagnostics. The F1-Score balances precision and recall, while PR-AUC emphasizes sensitivity across different thresholds [35]. The Matthews correlation coefficient (MCC) considers all outcomes from the confusion matrix, and Specificity ensures that benign cases are not overtreated. We also analyzed learning curves for accuracy and loss to assess convergence and detect any underfitting or overfitting issues. Additionally, confusion matrices provided insight into per-class errors, especially among similar lesions. These analyses informed our strategies, including data augmentation, reweighting, and targeted sampling, all aimed at improving generalization and diagnostic accuracy.

IV. Result Analysis

A. Comparing the Performance of Models

Table I demonstrates that EFormer-EA outperforms all other models across all metrics, achieving a PR-AUC of 98.56 and an MCC of 96.15 on the BreakHis dataset, as well as a PR-AUC of 99.43 and an MCC of 97.02 on the BUSI dataset. While MaxViT and DeiT III perform competitively in terms of F1 score and PR-AUC, they consistently fall short by 1-3% in MCC, indicating that their predictions are less balanced. Swin v2 demonstrates moderate stability but underperforms in PR-AUC, especially on BreakHis. In contrast, EAT shows the lowest performance in both datasets, which suggests poorer adaptability to different modalities.

TABLE I. PERFORMANCE COMPARISON OF TRANSFORMER-BASED MODELS ON BOTH DATASETS.

BreakHis				
Model	F1	PR-AUC	MCC	Specificity
Swin v2	96.82	96.14	93.88	95.73
MaxViT	97.24	95.87	93.21	96.92
DeiT III	96.53	97.03	94.51	94.96
EAT	95.91	96.42	93.87	97.21
EFormer-EA	98.27	98.56	96.15	98.08
BUSI				
Model	F1	PR-AUC	MCC	Specificity
Swin v2	97.03	97.26	94.63	96.47
MaxViT	96.15	96.82	94.07	95.98
DeiT III	97.58	97.14	94.81	94.89
EAT	95.27	96.03	93.28	97.54
EFormer-EA	98.46	99.43	97.02	97.86

B. Performance Validation

Figure 4 shows the confusion matrices for the BreakHis and BUSI test sets, illustrating the model's classification performance. For BreakHis, the model accurately identified 576 benign and 592 malignant cases, with only 10 benign misclassified as malignant and 8 malignant misclassified as benign. This indicates a good balance between sensitivity and specificity despite the high variability in histopathology. In BUSI, the model's performance is almost flawless, correctly classifying 42 normal, 36 benign, and 44 malignant cases, with just one normal misclassified as benign and one malignant as normal, and no misclassification between benign and malignant.

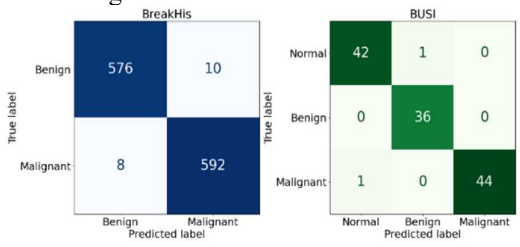

Fig. 4. Confusion Matrix of EFormer-EA on both datasets.

Figure 5 displays the training and validation loss and accuracy for BreakHis and BUSI over 30 epochs, illustrating smooth convergence and strong generalization. For BreakHis, both loss curves decline steadily with minimal separation, while accuracy consistently rises to over 98% by the final epoch. This indicates stable learning across the different magnifications of histopathology. In contrast, for BUSI, the loss values drop sharply in the early stages and then stabilize, while accuracy approaches nearly 99.34%.

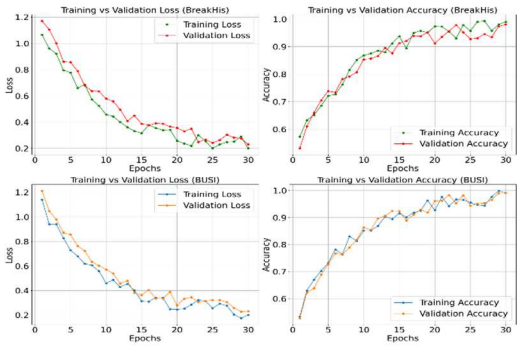

Fig. 5. Learning curve of EFormer-EA on both datasets.

C. State-of-the-art Comparison

Table II demonstrates that the proposed EFormer-EA model particularly enhances breast cancer identification through three key contributions. First, it achieves an impressive accuracy of 98.56% on the BreakHis dataset, surpassing the previous best of 98.10%, and reaches 99.43% on BUSI, outperforming CNN-U-Net by 13.43%. This showcases the effectiveness of its hybrid EfficientFormerV2–External Attention design in modeling both detailed histopathological features and broader ultrasound patterns. Second, in contrast to previous works ([19]– [34]), it incorporates Grad-CAM, which improves interpretability. Third, the proposed model has the potential for real-time application in a web-based platform.

TABLE II. COMPARISON WITH PREVIOUS MODELS.

Ref	Model	Data	Result	XAI + App
[19]	Hybrid	CBIS-DDSM, INbreast	98.00%	No
[21]	CNN-SOM	Breast MRI	98.00%	No
[22]	CNN-U-Net	BUSI	86.00%	No
[23]	CNN	INbreast	96.50%	No
[25]	ViT	IDC	92.12%	No
[26]	Hybrid	BreaKHis	98.10%	No
Ours	Eformer-EA.	BreaKHis	98.56%	Yes
Ours	Eformer-EA.	BUSI	99.43%	Yes

D. Explainable Web Application

Figure 6 shows the output of the web interface for an ultrasound image from the BUSI dataset. The left panel presents the original scan with a Grad-CAM heatmap, highlighting areas of focus with red and yellow for high relevance (on the lesion) and blue and green for lower relevance. The right panel lists class probabilities: Normal (0.79%), Benign (0.58%), and Malignant (98.63%), indicating strong confidence in a malignant diagnosis. This interface is designed for clinicians, offering real-time visualization and clear predictions to enhance trust and support point-of-care diagnostics.

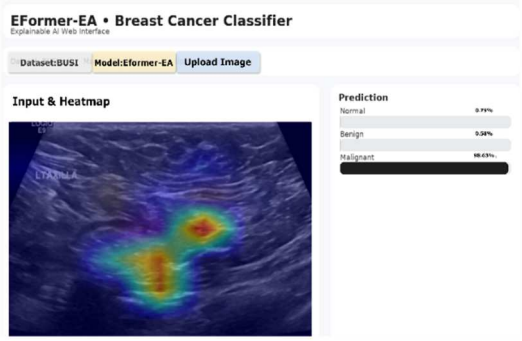

Fig. 6. Explainable web application.

V. DISCUSSION

EFormer-EA has outperformed all baseline models on the BreakHis and BUSI datasets, thanks to EfficientFormerV2's texture detail capture and External Attention's global dependency retention. Our preprocessing steps enhanced lesion contrast and boundary definition, leading to improved feature discrimination. GPU-accelerated augmentation effectively reduced overfitting and increased robustness against intra-class variations. These results are significant for computer-aided diagnosis in hospitals, labs, and telemedicine. The web interface provides real-time, interpretable predictions, facilitating remote case verification in resource-limited areas and can be adapted for veterinary tumor detection, aiding earlier diagnoses to reduce mortality and optimize treatment. Despite these, the model's performance is sensitive to external memory size and requires careful tuning, particularly at higher resolutions. While it is more efficient than many transformers, it still demands more resources than lightweight CNNs, which may limit its use in ultra-low-power scenarios. The evaluation was limited to two public datasets, requiring broader validation in multi-center clinical settings. Future work will concentrate on large-scale evaluations, optimization for edge devices, integration of multimodal clinical data, and exploring federated learning for privacy-preserving training across diverse datasets to enhance generalization and reliability.

VI. CONCLUSION

This study presents EFormer-EA, a hybrid transformer architecture that combines EfficientFormerV2 stages for effective local feature extraction and External Attention for global dependency capture. It focuses on multi-modality breast cancer imaging, addressing challenges like class imbalance, computational efficiency, and clinical interpretability. Key improvements include modality-specific preprocessing, histogram equalization, and GPU-accelerated augmentation to enhance feature quality. Class-weighted loss techniques also boosted performance for minority classes. Evaluations on BreakHis and BUSI achieved state-of-the-art results, showing significant improvements in PR-AUC and MCC compared to other transformer models. Additionally, Grad-CAM-based explainability was integrated into a web application to support clinicians. EFormer-EA stands out for its accuracy, efficiency, and interpretability, setting the stage for scalable AI-assisted breast cancer diagnosis. Future plans include large-scale clinical validation, optimization for edge devices, and integration with privacy-preserving training methods like federated learning.

REFERENCES

[1] H. Mewada, "Extended Deep-Learning Network for Histopathological Image-Based Multiclass Breast Cancer Classification Using Residual Features," Symmetry 2024, Vol. 16, Page 507, vol. 16, no. 5, p. 507, Apr. 2024, doi: 10.3390/SYM16050507.

[2] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," Biomed Signal Process Control, vol. 87, p. 105377, Jan. 2024, doi: 10.1016/J.BSPC.2023.105377.

[3] I. Nissar, S. Alam, and S. Masood, "Computationally efficient LC-SCS deep learning model for breast cancer classification using thermal imaging," Neural Comput Appl, vol. 36, no. 26, pp. 16233–16250, Sep. 2024, doi: 10.1007/S00521-024-09968-5/METRICS.

[4] R. Haque, S. H. Laskar, K. G. Khushbu, M. J. Hasan, and J. Uddin, "Data-Driven Solution to Identify Sentiments from Online Drug

Reviews," Computers 2023, Vol. 12, Page 87, vol. 12, no. 4, p. 87, Apr. 2023, doi: 10.3390/COMPUTERS12040087.

[5] N. Thakur, P. Kumar, and A. Kumar, "A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities," Multimedia Tools and Applications 2023 83:12, vol. 83, no. 12, pp. 35849–35942, Sep. 2023, doi: 10.1007/S11042-023-16634-W.

[6] M. Hayat, N. Ahmad, A. Nasir, and Z. Ahmad Tariq, "Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification," IEEE Access, vol. 12, pp. 184119–184131, 2024, doi: 10.1109/ACCESS.2024.3503413.

[7] F. Z. Nakach, A. Idri, and E. Goceri, "A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification," Artif Intell Rev, vol. 57, no. 12, pp. 1–53, Dec. 2024, doi: 10.1007/S10462-024-10984-Z/TABLES/3.

[8] M. M. Rahman, M. Najmul Gony, M. M. Rahman, M. M. Rahman, and S. D. Maria Khatun Shuvra, "Natural language processing in legal document analysis software: A systematic review of current approaches, challenges, and opportunities," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 3, pp. 5026–5042, Jun. 2025, doi: 10.53894/IJIRSS.V8I3.7702.

[9] M. M. Rahman, K. Dhakal, N. Gony, M. K. Shuvra, and M. Rahman, "AI integration in cybersecurity software: Threat detection and response," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 3, pp. 3907–3921, May 2025, doi: 10.53894/IJIRSS.V8I3.7403.

[10] M. Asaduzzaman, K. Dhakal, M. M. Rahman, M. M. Rahman, and S. Nahar, "Optimizing Indoor Positioning in Large Environments: AI," Journal of Information Systems Engineering and Management, vol. 10, no. 48s, pp. 254–260, May 2025, doi: 10.52783/JISEM.V10I48S.9500.

[11] S. M. Abid, Q. Xiaoping, M. M. Islam, M. A. Islam, M. M. Rahman, and A. R. M. Alam, "Edge-Conditioned GAT for Journal Ranking in Citation Networks," 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2025, pp. 1612–1619, 2025, doi: 10.1109/AINIT65432.2025.11035687.

[12] A. Pokharel, R. Dantu, S. Zaman, S. Talapuru, and V. Quach, "Deliberation Leads to Unanimous Consensus," 2024 6th International Conference on Blockchain Computing and Applications, BCCA 2024, pp. 356–363, 2024, doi: 10.1109/BCCA62388.2024.10844448.

[13] S. Zaman, R. Dantu, A. Pokharel, S. Talapuru, and V. Quach, "Indivisible State Mirroring in Cross-Chain Atomic Swap across Heterogeneous Blockchains," 2024 6th International Conference on Blockchain Computing and Applications, BCCA 2024, pp. 116–123, 2024, doi: 10.1109/BCCA62388.2024.10844452.

[14] S. Zaman, R. Dantu, S. Badruddoja, S. Talapuru, and K. Upadhyay, "Seamless Asset Exchange in Interconnected Metaverses: Unraveling On-Chain Atomic Swap," Proceedings - 2023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2023, pp. 146–155, 2023, doi: 10.1109/TPS-ISA58951.2023.00026.

[15] M. Sabbir Hossain et al., "Automatic Navigation and Self-Driving Technology in Agricultural Machinery: A State-of-the-Art Systematic Review," IEEE Access, vol. 13, pp. 94370–94401, 2025, doi: 10.1109/ACCESS.2025.3573324.

[16] M. S. Hossain, M. Rahman, M. Ahmed, M. M. Kabir, M. F. Mridha, and J. Shin, "DeepLabv3Att: Integrating Attention Mechanisms in DeepLabv3 for Enhanced Road Segmentation," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 711–718, 2024, doi: 10.1109/3ICT64318.2024.10824285.

[17] M. Rahman, M. S. Hossain, A. A. Eva, M. M. Kabir, M. F. Mridha, and J. Shin, "Alzheimers Disease Classification with a Hybrid CNN-SVM Approach on Enhanced MRI Data," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 50–57, 2024, doi: 10.1109/3ICT64318.2024.10824633.

[18] M. Rahman, M. S. Hossain, U. Rozario, S. Roy, M. F. Mridha, and N. Dey, "MultiSenseNet: Multi-Modal Deep Learning for Machine Failure Risk Prediction," IEEE Access, vol. 13, pp. 120404–120416, 2025, doi: 10.1109/ACCESS.2025.3586978.

[19] S. Chakravarthy et al., "Multi-class Breast Cancer Classification Using CNN Features Hybridization," International Journal of Computational Intelligence Systems, vol. 17, no. 1, pp. 1–19, Dec. 2024, doi: 10.1007/S44196-024-00593-7/TABLES/9.

[20] N. A. Wani, R. Kumar, and J. Bedi, "Harnessing Fusion Modeling for Enhanced Breast Cancer Classification through Interpretable Artificial Intelligence and In-Depth Explanations," Eng Appl Artif Intell, vol. 136, p. 108939, Oct. 2024, doi: 10.1016/J.ENGAPPAI.2024.108939.

[21] N. Ajlouni, A. Özyavaş, F. Ajlouni, M. Takaoğlu, and F. Takaoğlu, "Enhancing TNM Staging in Breast Cancer: A Hybrid Approach with CNN, Edge Detection, and Self-Organizing Maps for Improved Accuracy," 2024, doi: 10.21203/rs.3.rs-4794714/v1.

[22] B. Ş. Munteanu, A. Murariu, M. Nichitean, L. G. Pitac, and L. Dioşan, "Value of Original and Generated Ultrasound Data Towards Training Robust Classifiers for Breast Cancer Identification," Information Systems Frontiers, vol. 27, no. 1, pp. 75–96, Feb. 2025, doi: 10.1007/S10796-024-10499-6/TABLES/8.

[23] K. Jabeen, M. A. Khan, M. A. Hameed, O. Alqahtani, M. T. H. Alouane, and A. Masood, "A novel fusion framework of deep bottleneck residual convolutional neural network for breast cancer classification from mammogram images," Front Oncol, vol. 14, p. 1347856, 2024, doi: 10.3389/FONC.2024.1347856.

[24] M. M. Emam, E. H. Houssein, N. A. Samee, M. A. Alohali, and M. E. Hosney, "Breast cancer diagnosis using optimized deep convolutional neural network based on transfer learning technique and improved Coati optimization algorithm," Expert Syst Appl, vol. 255, p. 124581, Dec. 2024, doi: 10.1016/J.ESWA.2024.124581.

[25] A. Sriwastawa and J. A. Arul Jothi, "Vision transformer and its variants for image classification in digital breast cancer histopathology: a comparative study," Multimed Tools Appl, vol. 83, no. 13, pp. 39731–39753, Apr. 2024, doi: 10.1007/S11042-023-16954-X/METRICS.

[26] M. Hayat, N. Ahmad, A. Nasir, and Z. Ahmad Tariq, "Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification," IEEE Access, vol. 12, pp. 184119–184131, 2024, doi: 10.1109/ACCESS.2024.3503413.

[27] M. Shiri, M. P. Reddy, and J. Sun, "Supervised Contrastive Vision Transformer for Breast Histopathological Image Classification," Proceedings - 2024 IEEE International Conference on Information Reuse and Integration for Data Science, IRI 2024, pp. 296–301, 2024, doi: 10.1109/IRI62200.2024.00067.

[28] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," Proceedings of the International Joint Conference on Neural Networks, vol. 2016-October, pp. 2560–2567, Oct. 2016, doi: 10.1109/IJCNN.2016.7727519.

[29] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data Brief, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/J.DIB.2019.104863.

[30] Md. R. Ahmed et al., "Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis," Bioengineering 2025, Vol. 12, Page 651, vol. 12, no. 6, p. 651, Jun. 2025, doi: 10.3390/BIOENGINEERING12060651.

[31] J. Debnath et al., "LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis," Inform Med Unlocked, vol. 57, p. 101669, Jan. 2025, doi: 10.1016/J.IMU.2025.101669.

[32] A. S. U. K. Pranta et al., "A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification," Computers 2025, Vol. 14, Page 197, vol. 14, no. 5, p. 197, May 2025, doi: 10.3390/COMPUTERS14050197.

[33] R. Haque et al., "Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning," BioMedInformatics 2024, Vol. 4, Pages 966-991, vol. 4, no. 2, pp. 966–991, Apr. 2024, doi: 10.3390/BIOMEDINFORMATICS4020054.

[34] R. Haque et al., "Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis," Comput Biol Med, vol. 191, p. 110166, Jun. 2025, doi: 10.1016/J.COMPBIOMED.2025.110166.

[35] M. I. H. Siddiqui et al., "Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI," Inform Med Unlocked, vol. 56, p. 101657, Jan. 2025, doi: 10.1016/J.IMU.2025.101657.