ConvNeXt for Breast Cancer HER2 Scoring Using Different Types of Histopathological Stained Images

Lamiaa Abdel-Hamid

Electronics & Communication Department, Faculty of Engineering, Misr International University (MIU), Cairo, Egypt
Email: Lamiaa.a.hamid@miuegypt.edu.eg

***Abstract*—Human Epidermal Growth factor receptor 2 (HER2)-positive breast cancer is the most aggressive subtype, requiring targeted treatments for its effective management. Immunohistochemistry (IHC) is the gold standard for manual HER2 scoring that requires specialized antibodies and advanced lab equipment for accurate evaluation. Manual inspection of the IHC images requires a high level medical expertise, while having the downside of being tedious and extremely time consuming. Hematoxylin and Eosin (H&E) staining is the routine procedure for breast cancer detection making it more widely available and cost effective than IHC images. Deep learning-based methods can provide fast, reliable, and cost-efficient automated tools for HER2 scoring using histopathological images. ConvNeXt is a purely convolutional neural network that is based on ResNet and incorporates several advanced techniques inspired by vision transformers to enhance its performance. In this work, ConvNeXt is compared to three standard networks: InceptionV3, ResNet50, and MobileNetV2. The Breast Cancer Immunohistochemical (BCI) public dataset consisting of over three thousand IHC and H&E image pairs was used to evaluate the pretrained networks' performance. For both IHC and H&E images, ConvNeXt and ResNet50 achieved the highest accuracies. For 4-class HER2 classification, ConvNeXt attained accuracies of 97.79% and 95.58% for IHC and H&E images, respectively. These results outperform state-of-the-art methods from literature by up to 10%. Both IHC and H&E stained images are shown to be reliable for HER2-scoring using deep learning-based approaches. H&E stained images, given their low cost and widespread availability, thus represent strong candidates for integration into practical AI-assisted HER2 scoring systems.**

***Keywords*—breast cancer, Human Epidermal Growth factor receptor 2 (HER2) scoring, histopathological images, classification, Transfer Learning (TL), ConvNeXt**

I. INTRODUCTION

Breast cancer is the second most prevalent cancer worldwide, predominantly affecting woman. In 2020, there were over 2.3 million new cases of breast cancer, a number expected to exceed 3 million by 2040 [1]. Breast cancer risk factors include age, obesity, excessive alcohol and/or tobacco consumption, history of radiation exposure, reproductive history, postmenopausal hormone therapy, and family history of the disease. Unfortunately, even if all modifiable risk factors are controlled, the risk of developing breast cancer is only reduced by at most 30% [2]. Breast cancer can be either in situ or invasive. In situ breast cancer cells typically remain confined within the breast without spreading to the surrounding areas. Invasive breast cancer is more aggressive and can spread to other parts of the body such as the lungs, liver, bone, or brain. Regular screening can help detect breast cancer at an early stage when treatment is most likely to be successful.

Breast cancer screening typically involves undergoing an ultrasound and/or mammogram every one or two years, depending on the estimated risk factor. If a suspicious area is detected, a biopsy is requested by the healthcare provider to confirm or rule out the imaging results. A biopsy is a procedure in which a tissue sample is collected from the suspicious area for examination under a microscope to determine whether a suspicious area is benign or malignant (cancerous). Tissue samples are then stained using one or more staining materials to enhance the contrast of structures relevant to the disease. Staining helps pathologists better visualize the relevant tissue structures under the microscope, thereby facilitating diagnosis and increasing its reliability. The main staining methods used in cancer tissue analysis are: (1) Hematoxylin and Eosin (H&E) and (2) Immunohistochemistry (IHC).

H&E is the gold standard staining method used to determine whether a sample tissue is benign or malignant. Once breast cancer is confirmed, further tests are performed to check for the overexpression of specific proteins based on which breast cancer can be categorized into Human Epidermal Growth factor receptor 2 (HER2)-positive, Progesterone Receptor (PR)-positive, Estrogen Receptor (ER)-positive, or triple-negative if none of these proteins is detected [3, 4]. Identifying the type of breast cancer is essential as the treatment plan depends on the specific protein detected. HER2-positive breast cancer is the most aggressive type of breast cancer, as it is characterized by the rapid multiplication of cancer cells. HER2 targeted medicines work by blocking the cell growth signals, thus slowing down or completely stopping

Manuscript received February 17, 2025; revised March 18, 2025; accepted April 7, 2025; published July 15, 2025.

the growth of cancer cells [5, 6]. Current guidelines recommend that all invasive breast cancers be tested for the presence of the HER2 protein to determine whether HER2-targeted therapy, such as trastuzumab, is required.

IHC is a specialized staining method primarily used to determine the HER2 status of cancerous samples. The American Society for Clinical Oncology and the College of American Pathologists (ASCO/CAP) guidelines state that HER2 scoring from IHC images should be based on several parameters including staining intensity, completeness of the stained cell membrane, and percentage of positive cells [7]. Based on the ASCO/CAP guidelines, breast cancer tumors can be categorized into HER2-negative (scores 0 and 1+), borderline (score 2+), and HER2-positive (score 3+). Manual evaluation of IHC images by pathologists requires a high level of expertise, in addition to being tedious and time-consuming. Moreover, advanced lab equipment is necessary for IHC staining which may not be readily available in rural or underserved regions [8]. Deep learning-based methods have the potential to overcome these challenges by providing an automated and reliable AI-assisted HER2 scoring tool.

ConvNeXt was introduced by researchers from Facebook AI Research (FAIR) and UC Berkeley to compete with the rapidly emerging vision transformers. It is considered a modernized version of traditional Convolutional Neural Networks (CNNs) that incorporates several advanced deep learning techniques to enhance performance, such as such as depthwise separable convolutions, inverted bottleneck blocks, and advanced activation functions. In this work, ConvNeXt was compared to three standard convolutional networks for HER2 multiclass classification: InceptionV3, ResNet50, and MobileNetV2. IHC and H&E images were used as inputs to the pretrained networks to compare their usefulness for deep learning-based HER2 scoring. Several advanced deep learning techniques were applied to increase the robustness of the networks' training including data augmentation and regularization. Finally, results were compared to several state-of-the-art methods from the literature.

II. Literature Review

Deep learning methods have the advantage of learning both prominent and subtle relevant features from the data in an end-to-end manner. Standard networks are well-established models that have been previously developed and vigorously tested by the machine learning community. Popular examples include VGG, Inception, ResNet, and MobileNet families. Transfer Learning (TL) is a machine learning technique in which a model pre-trained on a task with abundant data is fine-tuned for a new task with limited data. Standard networks typically give good results for a wide range of tasks, especially when they are pretrained on large datasets. ImageNet [9] is a large-scale generic dataset that is widely used in the pretraining standard networks.

HER2 scoring methods typically rely on TL to achieve reliable performance. Tewary and Mukhopadhyay [10] compared five standard networks pretrained in ImageNet for 3-class HER2 classification: VGG16, VGG19, ResNet50, MobileNetV2, and NASNetMobile. Their experiments showed that the VGG19 network resulted in the highest accuracy of 93%, that was increased to 98% when an image-based statistical voting mechanism was considered. Wang *et al.* [11] presented a customized HER2-ResNet network that incorporated two different residual units along with convolutional filters having different kernel sizes, batch normalization, and pooling. For 4-class classification, Her2-ResNet resulted in an accuracy of 93% as compared to 89% and 75% for AlexNet and VGG16, respectively. Both summarized methods only considered IHC images.

Although H&E images are not currently used for manual HER2 scoring, several recent studies have demonstrated that deep learning models can achieve reliable performance when using H&E images. Wang *et al.* [12] introduced HAHNet that was based on multi-scale feature fusion using Inception along with channel attention modules. They compared their proposed network to seven standard networks for 4-class HER2 scoring using H&E images. HAHNet achieved an accuracy of 93.65% whereas the accuracies of the other networks ranged between 66% and 84%. Shovon *et al.* [5] presented HE-HER2Net which was based on the Xception network with a modified top layer including additional global average pooling, dropout, batch normalization, and dense layers with the Swish activation function. For 4-class HER2 classification, HE-HER2Net achieved an accuracy of 87% using H&E images, by that outperforming several other pretrained networks including DenseNet201, InceptionV3, EfficientNetB0, and VGG19. Senapati *et al*. [13] fused three different standard models for 4-class HER2 classification using H&E images: GoogLeNet, DenseNet201, and WideResNet50. Their model resulted in an accuracy of 97.84%, significantly outperforming previous methods. Nevertheless, ensembling deep networks comes at the cost of increased computational complexity. None of the previously mentioned methods compared the performance of deep learning-based HER2 scoring methods considering H&E and IHC images.

Mridha *et al*. [14] is among the few works that compared IHC and H&E images for 4-class HER2 classification. They presented the covoHER2 network, which was based on an InceptionV3 backbone with several added batch normalization and dense layers. IHC images resulted in an accuracy of 87.79% as opposed to an accuracy of 85.10% for the H&E images. Their results demonstrate that H&E images have good potential for deep learning-based HER2 scoring, achieving results comparable to those obtained with IHC images. However, they only tested their approach using the covoHER2 network and did not compare its performance to other standard networks.

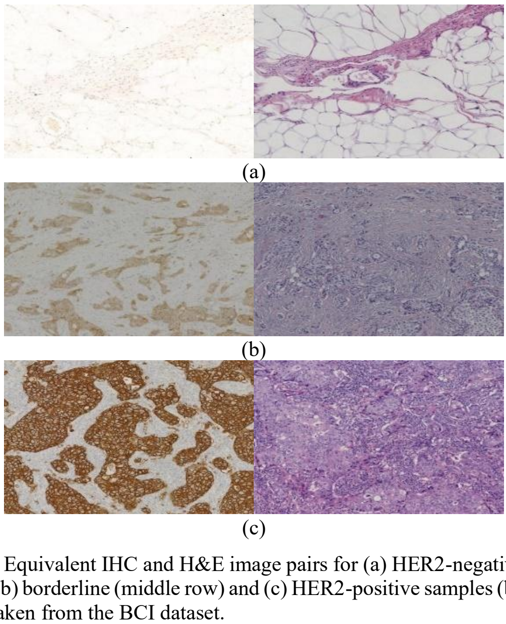
In this work, IHC and H&E images are compared for HER2 scoring considering four different pretrained standard networks: ConvNeXt, MobileNetV2, ResNet50, and InceptionV3. Data augmentation was employed to increase the diversity of the training dataset, thereby enhancing generalization. Additionally, a learning rate scheduler was utilized to achieve faster convergence during training. Experiments were performed using a public dataset that includes IHC and H&E image pairs, thus ensuring a fair comparison of the models' performance for the two staining methods.

III. Material and Methods

A. Dataset

Breast Cancer Immunohistochemical (BCI) dataset [15] includes equivalent IHC and H&E image patches taken from the whole slide image pairs of 51 breast cancer patients. The train and test datasets include images from different patients to ensure independence between them, resulting in more robust performance assessment [16].

In this work, dataset cleaning was performed to remove noisy images that contained no relevant cancer structures as was previously performed in [17, 18]. The BCI dataset utilized in the present study comprises 3,421 IHC/H&E image pairs with details including 1,101 HER2-negative images (score 0: 195 images, score 1+: 906 images), 1,606 borderline cases (score 2+), and 715 HER2-positive images (score 3+). Fig. 1 shows sample image pairs from these three categories. Approximately 90% of the total images were allocated to the training dataset. All images originally had a resolution of 1024×1024 pixels but were rescaled to 224×224 pixels to meet the input requirements of deep neural networks.



(a)

(b)

(c)

Fig. 1. Equivalent IHC and H&E image pairs for (a) HER2-negative (top row), (b) borderline (middle row) and (c) HER2-positive samples (bottom row) taken from the BCI dataset.

B. Standard Convolutional Neural Networks

In this work, four different standard networks pretrained on ImageNet were compared for HER2 scoring using either IHC or H&E images: (1) InceptionV3, (2) ResNet50, (3) MobileNetV2, and (4) ConvNeXt (Tiny).

1) InceptionV3 [19]: is based on the inception block, which combines convolution filters of different sizes in parallel to effectively capture multi-scale features. InceptionV3 improves upon its predecessors by implementing factorized convolutions, in which 3×3 convolutional filters are stacked to achieve larger receptive fields. This technique enhances the network's capacity to learn more complex features through increased depth without significantly increasing the computational cost.

2) ResNet50 [20]: introduced residual blocks to address the problem of vanishing or exploding gradients that arises in very deep networks. A residual block is formed when the output of a layer is directly passed to another layer further off in the network by that skipping one or more intermediate layers. Skip connections allow gradients to flow directly through the network, enabling efficient training of deep architectures.

3) MobileNetV2 [21]: is a lightweight model designed for resource-constrained devices such as mobile phones. MobileNetV2 uses depthwise separable convolution, which splits the computation into two steps: depthwise convolution and pointwise convolution. Depthwise convolution applies a single convolutional filter per input channel, while pointwise convolution applies a 1x1 convolution across all the input channels. This approach significantly reduces the computational cost compared to the standard convolution operation, resulting in fewer parameters and faster training.

4) ConvNext (Tiny) [22]: is a convolutional model based on the ResNet network, that is further enhanced through implementing several advanced techniques. ConvNeXt uses a patching strategy inspired by vision transformers, applying a 4×4 convolution kernel with a stride of 4 to partition the input into non-overlapping patches that are passed to the network for feature extraction. ConvNeXt also incorporates several techniques adopted from MobileNetV2 to further enhance performance, including depthwise separable convolutions, inverted bottlenecks, Gaussian Error Linear Units (GELU) activations, and layer normalization.

For all networks, a new classification head was added to fine-tune the models for the specific task. This head included a global average pooling layer used to reduce the number of parameters, followed by two fully connected layers of sizes 512 and 256 each accompanied by a dropout layer. Dropout is a regularization technique that works by randomly setting activations to zero during training, thus helping prevent overfitting and enhancing the model's generalization capability. Finally, a softmax activation function was used in the final layer for multiclass classification.

C. Network Training

Data augmentation introduces a wide variety of image transformations to the training dataset, thus enhancing its generalization capability. In the present study, several data augmentation techniques were applied, including horizontal and vertical flips, as well as horizontal and vertical shifts. Data augmentation was performed using the ImageDataGenerator in TensorFlow [23].

For the four implemented standard networks, all layers were set to '*trainable*' to ensure efficient fine-tuning on the target histopathological images, which differ significantly from the ImageNet dataset on which the networks were originally pretrained. All networks were trained for 30 epochs using an adaptive momentum estimation (Adam) optimizer and a categorical cross-entropy loss function. A learning rate scheduler was implemented for more robust training. Learning rate scheduling allows deep networks to start training at relatively large values promoting faster convergence during the initial stages. At later training stages, the learning rate is subsequently reduced to fine-tune the model's parameters and prevent it from overshooting the optimal solution. In this work, the initial learning rate was empirically chosen to be 10^{-4} and was reduced by a factor of 10 after 15 epochs. Experiments were performed using a computer equipped with an Intel Core i7-11800H CPU @2.3 GHz, 16 GB RAM and a GeForce RTX3060 6GB GPU. Due to hardware constraints, a batch size of 16 was utilized in all experiments.

D. Performance Metrics

To evaluate the performance of the presented models, the following metrics were used:

1) Accuracy: measures the overall correctness of the model by calculating the proportion of correctly classified instances out of the total number of instances.

$$Acc. = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2) Precision: measures the proportion of correctly predicted positive instances among all instances predicted as positive.

$$Prec. = \frac{TP}{TP+FP} \quad (2)$$

3) Recall: measures the proportion of actual positive instances that were correctly identified by the model.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

where TP, TN, FP and FN are true positive, true negative, false positive, and false negative, respectively.

IV. RESULTS AND DISCUSSION

In this work, four standard networks were implemented for HER2 multiclass classification: ConvNeXt, MobileNetV2, ResNet50, and InceptionV3. Additionally, two types of histopathological images were considered as inputs to the deep networks to investigate and compare their reliability for AI-assisted HER2 scoring: IHC and H&E stained images.

Experiments were conducted in two phases. In the first phase, 3-class HER2 classification (HER2-positive, borderline, and HER2-negative) was performed to compare the performance of the four implemented networks. In the second phase, the best-performing networks were selected for 4-class classification, categorizing images into HER2 scores 0, 1+, 2+, and 3+. Finally, these models were compared to several state-of-the-art methods from literature.

A. Experimental Results (3-Classes)

Tables I and II summarize the performance metrics of the four pretrained networks, considering the IHE and H&E images, respectively. Additionally, Figs. 2–5 illustrate the training accuracy curves of the four implemented networks. Overall, HER2 scoring accuracies for the ConvNeXt, InceptionV3, ResNet50, and MobileNetV2 pretrained on ImageNet ranged from 95% to 98% for the IHC images and from 86% to 95% for the H&E images. Since IHC stained images more directly convey HER2 related structures compared to H&E images, they achieved more superior performance.

TABLE I. HER2 3-CLASS CLASSIFICATION PERFORMANCE (IHC IMAGES)

Network	Epoch	Acc. (%)	Prec. (%)	Recall (%)
InceptionV3	19	96.56	96.54	96.07
ResNet50	25	97.30	97.30	97.30
MobileNetV2	28	95.33	95.79	95.29
ConvNeXt (Tiny)	**18**	**97.79**	**97.79**	**97.79**

TABLE II. HER2 3-CLASS CLASSIFICATION PERFORMANCE (H&E IMAGES)

Network	Epoch	Acc. (%)	Prec. (%)	Recall (%)
InceptionV3	23	89.93	90.48	88.70
ResNet50	29	93.91	93.84	93.61
MobileNetV2	28	86.73	87.22	85.50
ConvNeXt (Tiny)	**30**	**94.59**	**94.59**	**94.59**

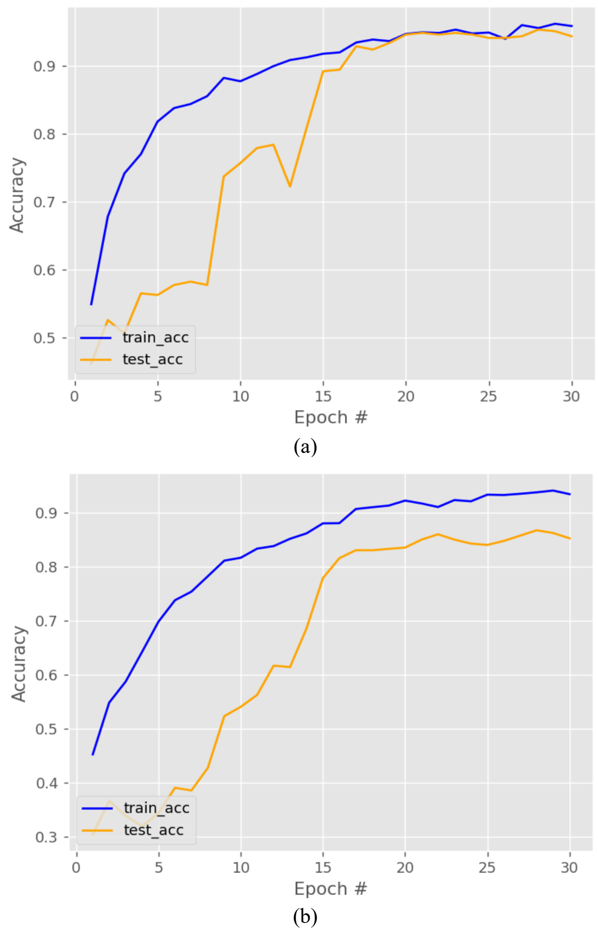
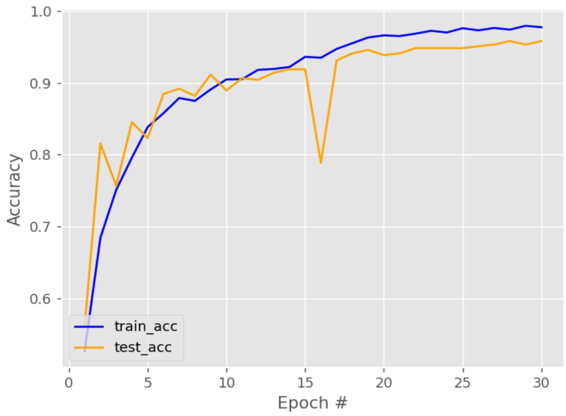
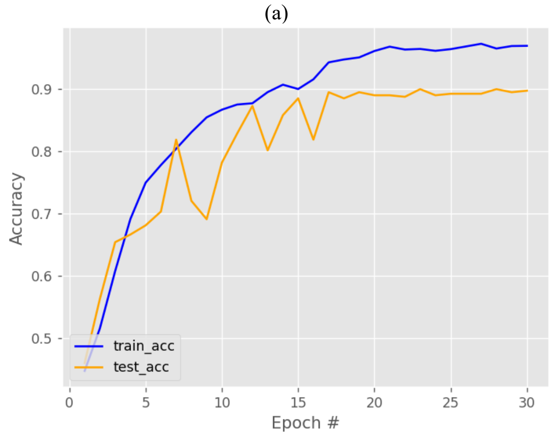


Fig. 2. MobileNetV2 training accuracy curves (a) IHC (b) H&E images.

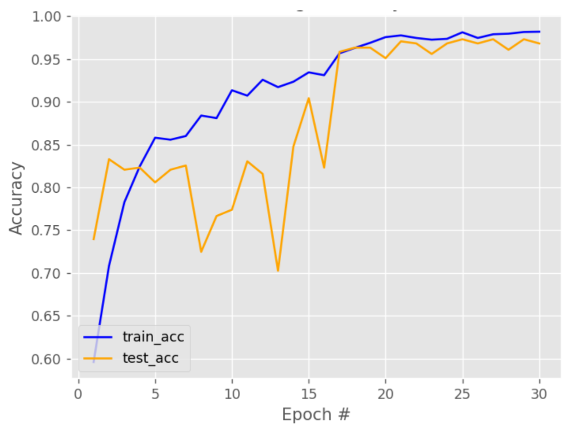


(a)

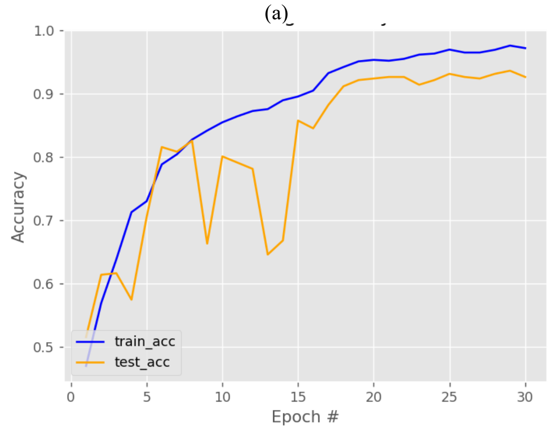


(b)

Fig. 3. InceptionV3 training accuracy curves (a) IHC (b) H&E images.

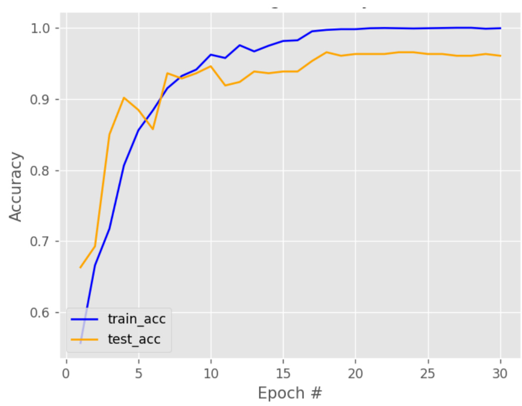


(a)

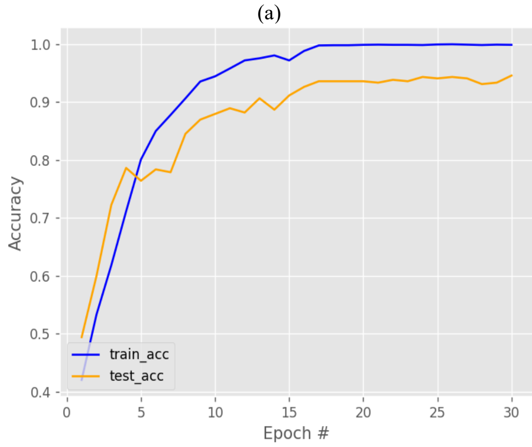


(b)

Fig. 4. ResNet50 training accuracy curves (a) IHC (b) H&E images.



(a)



(b)

Fig. 5. ConvNeXt (Tiny) training accuracy curves (a) IHC (b) H&E images.

For both types of histopathological images, the highest accuracies were achieved by ConvNeXt, closely followed by ResNet50. Specifically, ConvNeXt and ResNet outperformed the two other deep networks by 1.0–2.5% for the IHC images and 4.0–8.0% for H&E images. In previous work, Tewary and Mukhopadhyay [10] showed that VGG16 achieved an accuracy of 93% for 3-class HER2 classification using IHC images. In this work, ConvNeXt and ResNet50 resulted in accuracies of 97.30% and 97.79% for the IHC images, respectively. These results demonstrate the importance of considering advanced deep networks for challenging tasks, such as HER2 scoring, to be able to achieve reliable performance.

B. GradCam

Gradient-weighted Class Activation Mapping (Grad-CAM) is a widely used explainability technique that provides visual interpretations of a deep model's predictions, by highlighting the most relevant regions influencing its decision. In medical applications, explainability is crucial for building trust in AI-assisted diagnosis, as it allows clinicians to understand the reasoning behind model decisions. This enhances transparency, supports model validation, and facilitates integration into clinical workflows by providing interpretable insights for healthcare professionals.

Figs. 6 and 7 show the Grad-CAM heatmaps across the top convolutional layers in the three different ConvNeXt stages for sample IHC and H&E images, respectively. The generated heatmaps indicate that deeper convolutional layers are more effective at capturing meaningful structures within the histopathological images, whereas shallower layers capture little to no relevant information.

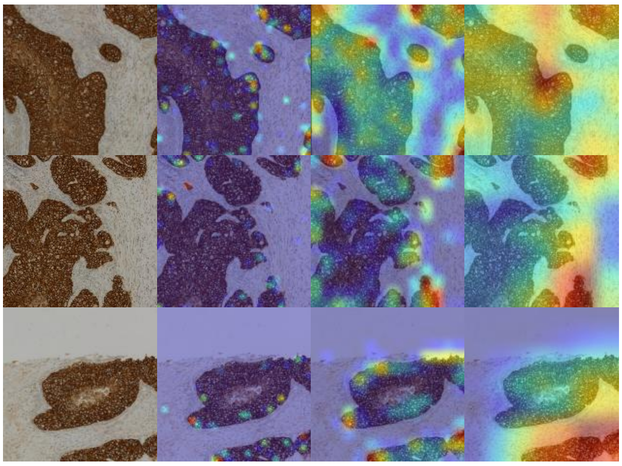


Fig. 6. Sample HER2-positive (score 3+) IHC images along with the generated Grad-CAM heatmaps for the top convolutional layers in the three different ConvNeXt stages.

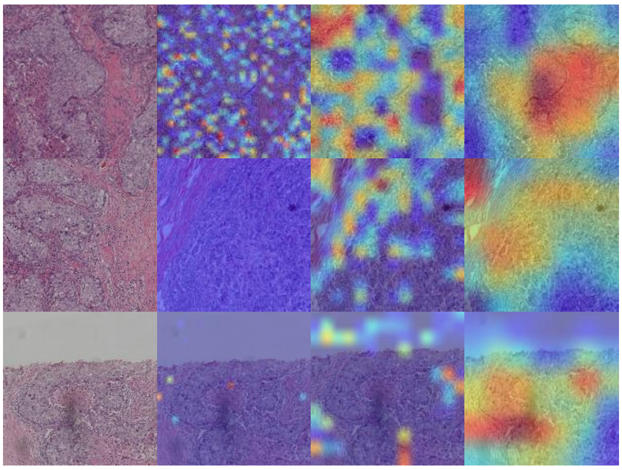


Fig. 7. Sample HER2-positive (score 3+) H&E image along with the generated Grad-CAM heatmaps for the top convolutional layers in the three different ConvNeXt stages.

C. Comparison to Previous Work (4-Classes)

In this subsection, 4-class HER2 classification was performed using the ConvNeXt and ResNet50 networks, as they achieved the highest accuracies in the previous subsection. Table III summarizes the performance metrics of the performed experiments, considering the IHC and H&E images. Both networks achieved accuracies ranging from 96% to 97.8% for the IHC images and from 95.5% to 96.6% for the H&E images. ConvNeXt and ResNet50 networks thus resulted in reliable performance regardless of whether the IHC or H&E images were considered.

Table IV summarizes the performance of several 4-class HER2 classification approaches from the literature in comparison to the ConvNeXt and ResNet50 models implemented in the present study. All the methods included in the comparison utilized the BCI dataset. ConvNeXt and ResNet50 outperformed most methods from literature by up to 10%. Senapati *et al.* [13] introduced an ensemble-based model in which three complex networks were fused to enhance HER2 classification performance using H&E images. Their fused model with class weighing resulted in an accuracy of 97.84%, which is slightly higher than the results obtained by the presented ConvNeXt network (95.58%). However, their approach involves a trade-off, as the integration of multiple complex deep networks significantly increases computational complexity and memory requirements.

TABLE III. CONVNEXT AND RESNET50 PERFORMANCE (4-CLASSES)

Network	Image Type	Acc. (%)	Prec. (%)	Recall (%)
ConvNeXt (Tiny)	IHC	97.79	97.79	97.79
	H&E	95.58	96.05	95.58
Resnet50	IHC	96.07	96.31	96.07
	H&E	96.56	96.55	96.31

TABLE IV. PERFORMANCE COMPARISON ON THE BCI DATASET (4-CLASSES)

Reference	Model	Image Type	Acc. (%)
Mridha *et al.* [14]	convoHER2 (based on InceptionV3)	IHC	87.79
		H&E	85.10
Shovon *et al.* [5]	HE-HER2Net (based on Xception)	H&E	87.01
Wang *et al.* [12]	HAHNet (based on InceptionV3)	H&E	93.65
Senapati *et al.* [13]	Fused GoogLeNet–DenseNet201–WideResNet50	H&E	97.84
Proposed	ResNet50	IHC	96.07
		H&E	96.56
	ConvNeXt (Tiny)	IHC	97.79
		H&E	95.58

V. CONCLUSIONS

Four state-of-the-art standard networks were implemented for HER2 scoring using histopathological images: InceptionV3, ResNet50, MobileNetV2, and ConvNeXt (Tiny). IHC and H&E stained images were compared to investigate their usefulness of the latter for HER2 scoring, despite being, so far, not used in manual HER2 scoring. Transfer learning was applied to achieve faster training and enhance performance. Several techniques were employed to ensure that the pretrained networks were effectively fine-tuned on the histopathological datasets including: (1) data augmentation to enhance the diversity of the training dataset, (2) addition of dropout layers in the newly added classification head to prevent overfitting, and (3) the use of a learning rate scheduler to facilitate training convergence. Experiments showed that ConvNeXt and ResNet50 were the best performing networks. For 4-class HER2 classification, ConvNeXt achieved accuracies of 97.79% and 95.58% for the IHC and H&E images, respectively. ResNet50 achieved accuracies of 96.07% and 96.56% for the IHC and H&E images, respectively. Both networks thus outperformed several state-of-the-art methods up to 10%.

IHC stained images are currently the gold standard for manual HER2 scoring. However, experiments conducted in this work demonstrate that AI-assisted HER2 scoring methods gave reliable performance for both IHC and H&E stained images. H&E staining is the standard technique used in routine pathology, making it more widely available and less expensive than IHC staining. Automated HER2 scoring methods that utilize H&E images would thus offer a cost effective and reliable tool than can be specifically useful in rural or underserved areas. Nevertheless, several clinical limitations remain unexplored in this work including its generalizability across more diverse datasets and the feasibility of integration within current pathology workflows. Future work will involve collaborating with breast cancer pathologists from specialized medical centers to further refine the presented AI-assisted HER grading method, ensuring its robustness and deployment readiness.

Conflict of Interest

The author declares no conflict of interest.

References

[1] M. Arnold *et al*., "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *Breast*, vol. 66, pp. 15–23, 2022. doi: 10.1016/j.breast.2022.08.010

[2] World Health Organization (WHO). Breast cancer. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[3] American Cancer Society: Breast Cancer Hormone Receptor Status. [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-hormone-receptor-status.html

[4] American Cancer Society: Breast Cancer HER2 Status. [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-her2-status.html

[5] M. S. H. Shovon, M. J. Islam, M. N. A. K. Nabil *et al*., "Strategies for enhancing the multi-stage classification performances of HER2 breast cancer from hematoxylin and eosin images," *Diagnostics*, vol. 12, no. 11, 2825, 2022. doi: 10.3390/diagnostics12112825

[6] Z. Xu, Q. Yang *et al*., "Predicting HER2 status in breast cancer on ultrasound images using deep learning method," *Front. Oncol*., vol. 12, 829041, 2022.

[7] G. Farshid *et al*., "ASCO/CAP 2018 breast cancer HER2 testing guidelines: Summary of pertinent recommendations for practice in Australia," *Pathology*, vol. 51, pp. 345–348, 2019. doi: 10.1016/j.pathol.2019.02.004

[8] Z. Han *et al*., "A deep learning quantification algorithm for HER2 scoring of gastric cancer," *Front. Neurosci*., vol. 16, 877229, 2022.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248–255.

[10] S. Tewary and S. Mukhopadhyay, "HER2 molecular marker scoring using transfer learning and decision level fusion," *J. Digit. Imaging*, vol. 34, pp. 667–677, 2021.

[11] X. Wang *et al*., "HER2-ResNet: A HER2 classification method based on deep residual network," *Technol. Health Care*, vol. 30, pp. 215–224, 2022.

[12] J. Wang, X. Zhu, K. Chen, L. Hao, and Y. Liu, "HAHNet: A convolutional neural network for HER2 status classification of breast cancer," *BMC Bioinformatics*, vol. 24, 353, 2023. doi: 10.1186/s12859-023-05474-y

[13] P. G. P, K. Senapati, and A. K. Pandey, "A novel decision level class-wise ensemble method in deep learning for automatic multi-class classification of HER2 breast cancer hematoxylin-eosin images," *IEEE Access*, vol. 12, pp. 46093–46103, 2024. doi: 10.1109/ACCESS.2024.3382212

[14] M. F. Mridha *et al*., "ConvoHER2: A deep neural network for multi-stage classification of HER2 breast cancer," arXiv preprint, arXiv:2211.10690, 2022

[15] S. Liu *et al*., "BCI: Breast cancer immunohistochemical image generation through pyramid pix2pix," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*., 2022, pp. 1815–1824.

[16] E. Conde-Sousa *et al*., "HEROHE Challenge: Predicting HER2 status in breast cancer from hematoxylin–eosin whole-slide imaging," *J. Imaging*, vol. 8, 213, 2022.

[17] C. Q. Cordeiro *et al*., "An automatic patch-based approach for HER-2 scoring in immunohistochemical breast cancer images using color features," arXiv preprint, arXiv:1805.05392, 2018

[18] D. Anand *et al*., "Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images," *J. Pathol. Inform*., vol. 11, 19, 2020.

[19] I. Ray *et al*., "Rethinking ImageNet pre-training for computational histopathology," in *Proc. 2022 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., 2022, pp. 3059–3062.

[20] K. He *et al*., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2016, pp. 770–778.

[21] M. Sandler *et al*., "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2018, pp. 4510–4520.

[22] Z. Liu *et al*., "A ConvNet for the 2020s," arXiv preprint, arXiv:2201.03545, 2022.

[23] TensorFlow: ImageDataGenerator. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator