TIPTEKNO'24 MEDICAL TECHNOLOGIES CONGRESS
10-12 October 2024
Yasmin Bodrum Resort, Muğla, Türkiye

Deep Learning Based Discovery of New Breast Cancer Stages: StyleGAN3 and Swin Transformer Approach

Reyhan Dede
Department of Computer Engineering
Yildiz Technical University (YTU)
Signal and Image Processing Lab. (SIMPLAB) at YTU
34220 Istanbul, Turkey
reyhan.dede@std.yildiz.edu.tr

Gokhan Bilgin
Department of Computer Engineering
Yildiz Technical University (YTU)
Signal and Image Processing Lab. (SIMPLAB) at YTU
34220 Istanbul, Turkey
gbilgin@yildiz.edu.tr

***Abstract*—This paper introduces a novel deep learning framework for the discovery of breast cancer stages, which integrates GAN-generated synthetic images with multi-omics data. By employing StyleGAN3 for the generation of realistic histopathological images and Swin Transformer for classification, the model draws upon both visual and biological data to enhance the accuracy of cancer staging predictions. The proposed methodology entails the generation of high-quality synthetic images using StyleGAN3, with a Fréchet Inception Distance (FID) score of 35, indicating a reasonable degree of similarity to real images. The images, in conjunction with RNA, miRNA, and clinical data, are integrated into a Swin Transformer-based classifier, resulting in an accuracy of 95.03%, a precision of 95.00%, and an F1 score of 95.00%. A threshold-based softmax probability analysis was employed during the inference stage to explore the potential discovery of new cancer stages. The preliminary observation-based threshold of 30% may be optimized through further experimentation. In the event that the model exhibited a confidence level for a given class below the specified threshold, the image was identified as a potential candidate for a previously unidentified stage. This study underscores the potential of multimodal data integration in enhancing breast cancer staging and offers insights into leveraging deep learning models for generating and classifying histopathological data, alongside identifying novel disease stages.**

***Keywords*—Breast cancer, StyleGAN3, Swin Transformer, Multimodal LLM, Cancer staging, New cancer stage discovery.**

I. Introduction

Breast cancer is one of the most prevalent forms of cancer among women globally, with millions of new cases diagnosed annually. In 2022, approximately 2.3 million women were diagnosed with breast cancer, representing the most prevalent form of cancer among women [1].

Molecular biomarkers are crucial for identifying breast cancer subtypes, particularly triple-negative breast cancer (TNBC), an aggressive form lacking ER, PR, and HER2 biomarkers. TNBC has limited response to current treatments, and a deeper understanding of its molecular distinctions is vital for developing more effective, personalized therapies.[2].

979-8-3315-2981-9/24 $31.00 © 2024 IEEE

Artificial intelligence (AI), particularly deep learning methods, has transformed the diagnosis and classification of breast cancer. Generative Adversarial Networks (GANs) have made a substantial impact on biomedical imaging by facilitating the generation of high-resolution synthetic data from histopathological images [3]. These networks facilitate the generation of data for rare cancer subtypes and enhance flexibility in clinical processes by producing realistic synthetic images from incomplete or limited data [4]. Despite the promise of GANs, many existing studies focus exclusively on image data, without considering the integration of complementary biological data, such as RNA, miRNA, and clinical records. These additional data types could provide a more complete understanding of cancer heterogeneity [5], [6].

To improve the accuracies of segmentation and classification results obtained by supervised and unsupervised learning methods, spatial information must also be considered in every approach [7], [8]. Recent studies have underscored the potential of multimodal data fusion, which entails integrating molecular, histopathological, radiological, and clinical data, to enhance the discovery and diagnosis of cancer biomarkers. These approaches permit a more comprehensive view of tumor biology, thereby facilitating the development of more personalized treatment strategies [9].

The field of oncology faces significant challenges, including data scarcity, standardization issues, and difficulties in interpreting multimodal models. While some studies have explored multimodal integration, the full potential of AI in oncology remains largely underutilized. Many models lack interpretability or clinical validation, limiting their clinical applicability. Although AI has shown promise in improving diagnostic accuracy in radiology, histopathology, and genomics, most studies do not adequately address its impact on cancer staging, revealing a gap in the literature on multimodal data fusion in cancer progression.[10].

Furthermore, hybrid architectures, such as those combining Vision Transformers (ViTs) with convolutional neural networks (CNNs), have recently been employed to enhance the

performance of models in open-set classification tasks [11]. These architectures have demonstrated enhanced accuracy in handling out-of-distribution samples and in improving the interpretability of AI models, particularly in scenarios involving the analysis of previously unseen data types or manipulations. Nevertheless, their application in the field of medical image analysis, particularly for the integration of biological data with histopathological images, remains constrained [11].

Recent advancements in generative models, such as GANs and diffusion models, have opened new avenues for synthesizing medical imaging data to address demographic representation biases. This aims to improve diagnostic models' performance across diverse populations, addressing imbalances in existing medical datasets. However, further research is needed to validate these approaches in clinical settings and investigate their efficacy across various medical conditions and demographic groups [12].

The objective of this study is to develop a comprehensive breast cancer staging model that incorporates both image and biological data. We utilize GANs to generate realistic synthetic histopathological images and integrate them with biological data, including RNA, miRNA, and clinical records, using advanced deep learning models such as the Swin Transformer [13]. By integrating these data modalities, our objective is to create a more comprehensive representation of tumor heterogeneity, thereby facilitating the discovery of new cancer stages and more personalized treatment strategies. Our approach is distinct from existing models in that it fuses multimodal data, addressing the limitations of single-modality models and improving the accuracy of breast cancer staging.

II. Methodology

A. Data Set

1) Dataset Overview: In this study, we used the CPTAC-BRCA dataset provided by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) on The Cancer Imaging Archive (TCIA) platform [14]. The dataset is widely used in breast cancer research and consists of 642 histopathological images in SVS format, along with biospecimen DNA, RNA, miRNA, WXS, and clinical data from 134 patients.
The dataset includes images categorized into 7 distinct breast cancer stages: Stage IA, Stage IIA, Stage IIB, Stage III, Stage IIIA, Stage IIIB, and Stage IIIC. However, there is a significant imbalance in class distribution, with Stage IIIC containing 4,200 samples, while the least represented stage, Stage IA, contains only 240 samples. Other stages similarly exhibit varying levels of representation, with the dataset containing a total of 9,065 images prior to any augmentation.

2) Data Preprocessing: During data preprocessing, the histopathological images were segmented into patches of 256x256 pixels. Low-resolution preview images were obtained from the SVS files, followed by binary thresholding to identify significant regions. Adaptive histogram equalization was applied to enhance contrast, yielding 19,515 image patches

Figure 1. We excluded 34 patients with missing cancer stage information and conducted the analysis on the remaining 100 patients with complete cancer stage data.

3) Data Augmentation and Balancing: To address the issue of class imbalance, data augmentation techniques were employed to ensure that all cancer stage classes contained an equal number of samples. Specifically, the sample size for each class was augmented in order to match the highest class size, which was 4,200 samples in Stage IIIC. Augmentation techniques, including random horizontal and vertical translations, rotations, and color adjustments, were employed to enhance the representation of the underrepresented classes. This resulted in a balanced dataset, with each class containing 4,200 samples. Following the augmentation process, the total number of images in the dataset was increased from 9,065 to 29,400, thereby ensuring that each class contributed equally to the training process.

4) Supporting Data Fusion: Biospecimen DNA, RNA, miRNA, WXS, and clinical data were converted into fixed-length tensors for each patient. These tensors were then fused and used as input data to effectively integrate biomolecular and clinical information into the model.

B. Model Training

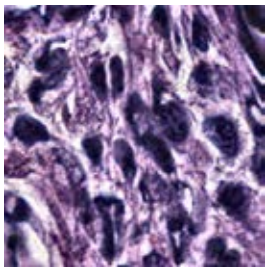


Fig. 1: Input Image

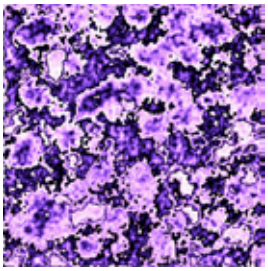


Fig. 2: Generated Image

1) Feature Extraction with Swin Transformer: The Swin Transformer was selected for its ability to capture both local and global visual information efficiently. The process begins by dividing the input image into non-overlapping windows, calculating attention within each window. The hierarchical structure allows the model to process large inputs while maintaining computational efficiency. To capture cross-window dependencies, a shifted window mechanism is used, shifting the windows after each block, allowing more comprehensive feature extraction across the entire image.

Additionally, the architectural design includes residual post-normalization, normalizing the outputs of each residual block before merging them back into the main branch. This helps prevent the accumulation of high activation amplitudes, stabilizing deeper layers. Attention is computed using a scaled cosine similarity function instead of the conventional dot product, mitigating issues related to extreme attention values and allowing more refined feature extraction.

To handle variations in input resolution, the model employs a log-spaced continuous position bias, enabling seamless interpolation of positional encodings across different window

sizes. This adaptability ensures that the Swin Transformer can maintain consistent performance across varying image scales without a significant drop in accuracy.

2) Training Process: The model's training relies on two main components: feature extraction with the Swin Transformer and image generation using a StyleGAN3-based GAN architecture. StyleGAN3 addresses aliasing issues by removing the dependence of fine details on pixel coordinates and ensuring full sub-pixel translation and rotation equivariance. It also uses high-quality filters and low-pass filtering to prevent unwanted information from propagating, resulting in more consistent and realistic image generation [15]. The Swin Transformer extracts hierarchical features from real histopathological images, which are combined with multi-omics data and fed into both the GAN's generator and discriminator. The generator synthesizes 256x256 images from random noise vectors and support data, while the discriminator evaluates their authenticity.

The discriminator is optimized using Binary Cross-Entropy (BCE) loss, and a perceptual loss function based on features from a pre-trained VGG19 network guides the generator to produce more detailed, realistic images. Both the generator and discriminator are trained with Stochastic Gradient Descent (SGD) optimization, and gradient clipping is applied to stabilize training. Early stopping is used to avoid overfitting, and a learning rate scheduler ensures smooth convergence. R1 regularization is applied to improve generalization by penalizing sharp decision boundaries. The process continues until early stopping criteria are met or the maximum number of epochs is reached.

C. Classification

Data Labeling and Stage Identification The study focused on classifying breast cancer stages and identifying potential new stages using deep learning models, including Swin Transformer, Vision Transformer (ViT), EfficientNet, ResNet, and Inception. The stages of the pre-processed images were identified by matching patient identities from clinical support data with histopathology images, ensuring accurate labeling. These stages were converted into numerical values via a label encoder for classification tasks.

Swin Transformer outperformed other models, achieving 92.24% accuracy and a 92.28% F1 score, due to its hierarchical structure, shifted window mechanism, and layer normalization, which enabled it to effectively learn both local and global features. Enhancements like Scaled Cosine Attention and residual post-norm contributed to balanced learning and stable training. ViT followed closely with an accuracy of 92.18% and F1 score of 92.07%, excelling in multi-scale feature learning, which is crucial for medical imaging.

EfficientNet performed well with 91.56% accuracy but lagged behind the Transformer-based models in handling the complexity of histopathology images. ResNet and Inception showed weaker results, particularly Inception, which struggled to generalize across cancer stages due to its reliance on fixed-size kernels.

To further enhance Swin Transformer, the dataset was expanded, and train-test-validation separation was reorganized. This increased data variety improved generalization, boosting accuracy to 95.03% and F1 score to 95.00%. Data augmentation techniques, including random flips and rotations, were applied to prevent overfitting, while early stopping ensured the model did not overtrain.

The study also utilized synthetic histopathology images generated by StyleGAN3. Swin Transformer successfully classified these images and identified potential new stages. Overall, Swin Transformer proved most effective, and the integration of synthetic images and additional training data further improved its performance, offering a more detailed understanding of breast cancer stages beyond current classification systems.

III. Evaluation

The proposed conditional GAN model, which integrates StyleGAN3 and Swin Transformer for breast cancer stage discovery, was evaluated using both qualitative and quantitative metrics. The quality of synthetic images was assessed through the Frechet Inception Distance (FID), while the performance of the Swin Transformer model was measured using standard metrics, including accuracy, precision, recall, and F1 score. Furthermore, new stage discovery was explored through a threshold-based softmax analysis.

A. Frechet Inception Distance (FID)

In this study, Frechet Inception Distance (FID) was employed to evaluate the quality of synthetic images generated by StyleGAN3. FID calculates the distance between the feature distributions of real and generated images, as extracted by the Inception-v3 model[16]. A FID score of **35** was achieved, indicating a reasonable level of similarity between the synthetic and real images. .

B. Swin Transformer Classification Performance

The model's strong accuracy and consistency across both real and synthetic images confirm the effectiveness of the GAN-generated data for breast cancer stage classification.

C. Confusion Matrix Analysis

Figure 3 shows the confusion matrix for the Swin Transformer model. Minor misclassifications occurred between visually similar stages (e.g., Stage IIA/IIB and Stage IIIB/IIIC), likely due to biological similarities. Incorporating omics data could improve differentiation in these cases.

D. Softmax Probability Analysis for New Stage Discovery

To identify potential new stages, we applied a threshold-based softmax analysis during inference. A threshold of **30%** was set, flagging images with low classification confidence as possible new stages. This approach can be further refined with additional validation to aid in the discovery of previously unidentified cancer stages.

IV. Limitations and Future Directions

The study shows that GAN-based methods and Swin Transformer can identify new breast cancer stages through histopathologic images and supplementary data. However, limitations exist, such as the lack of segmentation or spatial analysis of tumor regions. The threshold-based softmax analysis is still in its preliminary phase and requires further validation. Future work may incorporate multimodal large language models (LLM) to enhance staging accuracy, particularly in identifying new stages with greater detail by analyzing histopathological images and support data. LLMs can provide a more nuanced understanding of the clinical implications of new stages.

TABLE I: Performance metrics of different models and optimizers

Model	Optimizer	Accuracy	Precision	Recall	F1 Score	Learning Rate
Swin	Adam	0.1415	0.0200	0.1415	0.035	0.001
Swin	SGD	0.9224	0.9238	0.9224	0.9228	0.001
ResNet	Adam	0.7986	0.7996	0.7986	0.7963	0.001
ResNet	SGD	0.5463	0.5640	0.5463	0.5371	0.001
EfficientNet	Adam	0.8633	0.8621	0.8633	0.8610	0.001
EfficientNet	SGD	0.9156	0.9149	0.9156	0.9151	0.001
ViT	Adam	0.5014	0.5192	0.5014	0.4737	0.001
ViT	SGD	0.9218	0.9208	0.9218	0.9207	0.001
Inception	Adam	0.7054	0.7347	0.7054	0.6967	0.001
Inception	SGD	0.9190	0.9181	0.9190	0.9179	0.001
Swin	Adam	0.1388	0.0193	0.1388	0.0338	0.0005
Swin	SGD	0.9163	0.9201	0.9163	0.9164	0.0005
ResNet	Adam	0.8633	0.8694	0.8633	0.8625	0.0005
ResNet	SGD	0.4946	0.4777	0.4946	0.4734	0.0005
EfficientNet	Adam	0.9184	0.9200	0.9184	0.9171	0.0005
EfficientNet	SGD	0.8844	0.8834	0.8844	0.8824	0.0005
ViT	Adam	0.6469	0.6618	0.6469	0.6394	0.0005
ViT	SGD	0.9170	0.9205	0.9170	0.9170	0.0005
Inception	Adam	0.8184	0.8199	0.8184	0.8139	0.0005
Inception	SGD	0.8837	0.8845	0.8837	0.8804	0.0005

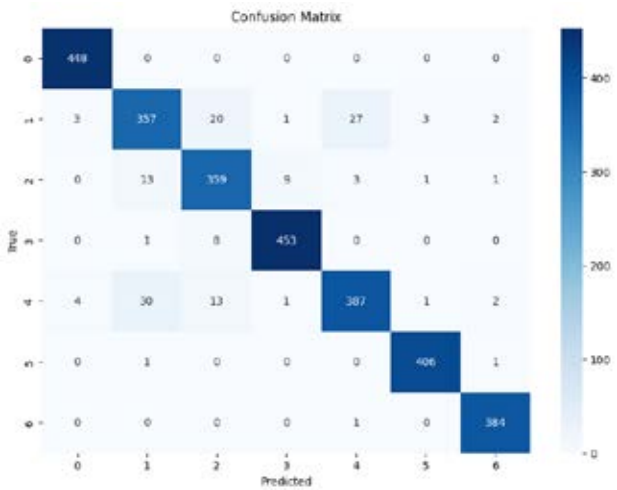


Fig. 3: Confusion matrix for the Swin transformer classifier on the updated dataset

TABLE II: Classification performance of Swin transformer

Metric	Value
Accuracy	0.9503
Precision	0.9500
Recall	0.9503
F1 Score	0.9500

References

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA: a cancer journal for clinicians*, vol. 73, no. 1, pp. 17–48, 2023.

[2] S. Zhu, Y. Wu, B. Song, M. Yi, Y. Yan, Q. Mei, and K. Wu, "Recent advances in targeted strategies for triple-negative breast cancer," *Journal of Hematology & Oncology*, vol. 16, no. 1, p. 100, 2023.

[3] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4217–4228, 2021.

[4] P. Antoniadis, C. Sartzetaki, N. Antonopoulos, P. Papageorgiou, A. Korfiati, and V. Pitsikalis, "Mask-cscgan for realistic synthetic cell generation," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 4575–4583.

[5] C. Zhu, X. Chai, Y. Xiao, X. Liu, R. Zhang, Z. Yang, and Z. Wang, "Swin-net: A swin-transformer-based network combing with multi-scale features for segmentation of breast tumor ultrasound images," *Diagnostics*, vol. 14, no. 3, p. 269, 2024.

[6] Z. Yao, T. Luo, Y. Dong, X. Jia, Y. Deng, G. Wu, Y. Zhu, J. Zhang, J. Liu, L. Yang *et al.*, "Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis," *Nature Communications*, vol. 14, no. 1, p. 788, 2023.

[7] G. Bilgin, "Evaluation of spatial relations in the segmentation of histopathological images," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2013, pp. 1–4.

[8] A. Albayrak and G. Bilgin, "Breast cancer mitosis detection in histopathological images with spatial feature extraction," in *Sixth International Conference on Machine Vision (ICMV 2013)*, vol. 9067. SPIE, 2013, pp. 103–108.

[9] S. Steyaert, M. Pizurica, D. Nagaraj, P. Khandelwal, T. Hernandez-Boussard, A. J. Gentles, and O. Gevaert, "Multimodal data fusion for cancer biomarker discovery with deep learning," *Nature machine intelligence*, vol. 5, no. 4, pp. 351–362, 2023.

[10] J. Lipkova, R. J. Chen, B. Chen, M. Y. Lu, M. Barbieri, D. Shao, A. J. Vaidya, C. Chen, L. Zhuang, D. F. Williamson *et al.*, "Artificial intelligence for multimodal data integration in oncology," *Cancer cell*, vol. 40, no. 10, pp. 1095–1110, 2022.

[11] J. Wang, O. Alamayreh, B. Tondi, and M. Barni, "Open set classification of gan-based image manipulations via a vit-based hybrid architecture," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 953–962.

[12] R. S. Fitzpatrick, "Using generative models to create synthetic medical imaging data to boost model performance on sparse demographic groups," Master's thesis, University of Hawai'i at Manoa, 2024.

[13] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, "Swin transformer v2: Scaling up capacity and resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 009–12 019.

[14] D. P. Clark, M. Gulley, M. Marko, J. Samayoa, R. T. Kelly, and S. Sinha, "CPTAC Breast Cancer (CPTAC-BRCA) Collection [Data set]," 2021, [Online]. Available: https://doi.org/10.7937/TCIA.2021.YI7CP4XS.

[15] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," *Advances in neural information processing systems*, vol. 34, pp. 852–863, 2021.

[16] S. Jayasumana, S. Ramalingam, A. Veit, D. Glasner, A. Chakrabarti, and S. Kumar, "Rethinking fid: Towards a better evaluation metric for image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 9307–9315.