BREAST CANCER DIAGNOSIS ON PATHOLOGICAL IMAGES USING DATA AUGMENTATION METHOD: CYCLE GAN

1 K N SANJANA (sanjanaraj668@gmail.com)

Guide Name: Dr.D.Aruna Kumari – Professor and HOD ,

VIDYA JYOTHI INSTITUTE OF TECHNOLOGY(An Autonomous Institution)

(Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad)

Aziz Nagar Gate, C.B. Post, Hyderabad-500075

ABSTRACT

Early and accurate diagnosis of breast cancer is critical for effective treatment and patient survival. This study explores the use of deep learning models for breast cancer classification using pathological images, enhanced through data augmentation using CycleGAN—a generative adversarial network capable of translating images between domains without paired training examples. The CycleGAN-based augmentation method helps in overcoming the limitations of small and imbalanced datasets, a common challenge in medical imaging.

In this research, several convolutional neural network architectures, including ResNet50, ResNet101, GoogleNet, VGG16, and Alex Net, were trained and evaluated on augmented pathological image datasets. Key performance metrics such as FScore, Recall, Precision, and Accuracy were calculated for each model to determine their effectiveness in breast cancer detection. Among the tested architectures, AlexNet achieved the highest overall performance with an accuracy of 88.76%, followed by VGG16 and GoogleNet. The results demonstrate that CycleGAN-based augmentation significantly improves model performance by increasing the diversity and quantity of training data, enabling better generalization and robustness in classification.

This work highlights the potential of combining generative augmentation techniques with deep learning for enhanced diagnostic accuracy in breast cancer pathology, supporting the development of more reliable computer-aided diagnosis systems.

1.INTRODUCTION

Breast cancer remains one of the most prevalent and life-threatening diseases among women worldwide. According to the World Health Organization (WHO), early detection and accurate diagnosis are crucial for improving survival rates and reducing treatment burdens. Histopathological image analysis, commonly used in breast cancer diagnosis, involves microscopic examination of tissue samples. However, manual interpretation by pathologists can be time-consuming, subjective, and prone to inter-observer variability. Therefore, there is a growing interest in automating this process using deep learning-based techniques, which have shown remarkable success in medical image classification tasks.

Deep convolutional neural networks (CNNs) such as ResNet, GoogleNet, VGG16, and AlexNet have been widely adopted for image-based classification due to their ability to learn complex hierarchical features. However, one major challenge in medical imaging is the limited availability of large, balanced datasets. Small sample sizes and class imbalance can hinder model performance and generalization. To address this limitation, data augmentation techniques are commonly employed. While traditional augmentation methods involve geometric and photometric transformations, they may not sufficiently capture the complex variations in pathological images.

In this study, we propose a more advanced augmentation approach using CycleGAN (Cycle-Consistent Generative Adversarial Network), a powerful deep learning model capable of generating realistic synthetic images without requiring paired data. CycleGAN learns to map images from one domain to another and has shown promising results in medical image synthesis. By applying CycleGAN, we aim to enrich the training dataset with high-quality synthetic images that reflect the intricate patterns of breast cancer pathology.

We evaluate the impact of CycleGAN-based augmentation on the performance of various CNN architectures—ResNet50, ResNet101, GoogleNet, VGG16, and AlexNet—in classifying breast cancer from histopathological images. The models are compared using standard performance metrics such as FScore, Recall, Precision, and Accuracy to determine the most effective approach for breast cancer detection.

This research demonstrates the potential of combining generative augmentation and deep learning for improving diagnostic accuracy, contributing to the development of more reliable and efficient computer-aided diagnosis (CAD) systems in the field of oncology.

II.RELATED WORK

The application of deep learning to medical image analysis, particularly for cancer diagnosis, has gained significant momentum in recent years. Convolutional Neural Networks (CNNs) have emerged as powerful tools for automated classification tasks, especially in histopathological imaging, due to their ability to learn complex features directly from raw pixel data.

Several studies have demonstrated the effectiveness of CNNs in breast cancer classification. For example, Spanhol et al. [1] introduced a benchmark dataset of breast histopathology images and evaluated the performance of CNNs in distinguishing benign from malignant tissues. Their work laid the foundation for future research in automated breast cancer detection using deep learning. Following this, researchers have explored deeper and more complex architectures like VGGNet, ResNet, and Inception (GoogleNet), which have shown superior performance in capturing intricate histological patterns.

Despite these advances, the availability of large, annotated medical datasets remains a major bottleneck. Small datasets not only limit the training of deep models but also introduce overfitting. To mitigate this, traditional data augmentation techniques such as rotation, flipping, zooming, and color jittering have been widely used to artificially increase dataset diversity. However, these transformations are often insufficient to capture the high variability and subtle morphological changes in histopathological images.

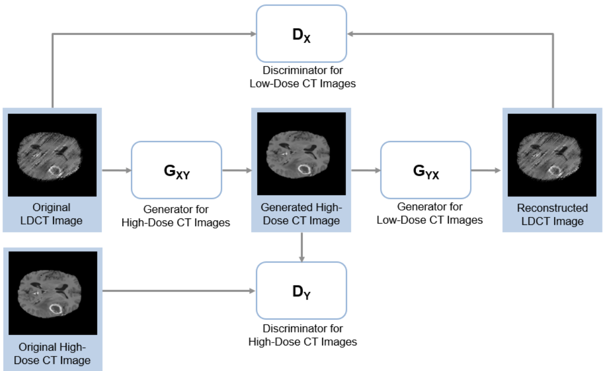
To address this limitation, recent studies have turned to **Generative Adversarial Networks (GANs)**. GANs have proven effective in synthesizing realistic medical images, thereby enriching training datasets. Zhu et al. [2] introduced **CycleGAN**, a variant of GAN capable of learning mappings between unpaired image domains. Its ability to preserve the structure while altering image style makes it particularly suitable for medical image augmentation.

In the context of breast cancer diagnosis, several researchers have applied GAN-based methods to improve classification accuracy. Baur et al. [3] used GANs for semantic augmentation of brain tumor MRI data, while Yi et al. [4] demonstrated the effectiveness of CycleGAN in translating between different histological staining styles to aid classification.

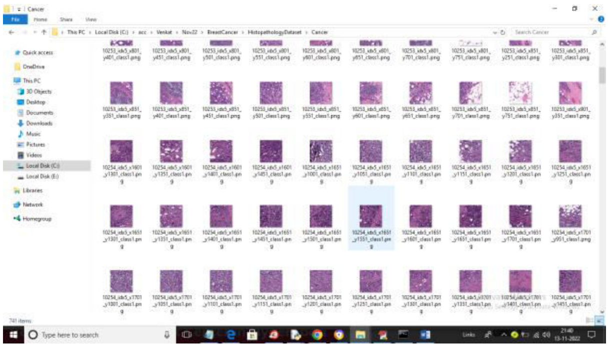
Despite these advancements, there remains limited work specifically integrating CycleGAN with multiple CNN classifiers to comprehensively evaluate performance gains in breast cancer histopathology. This study fills that gap by comparing the effects of CycleGAN-based augmentation on multiple well-known CNN architectures—including ResNet50, ResNet101, GoogleNet, VGG16, and AlexNet—providing a broader understanding of how generative augmentation impacts diagnostic performance.

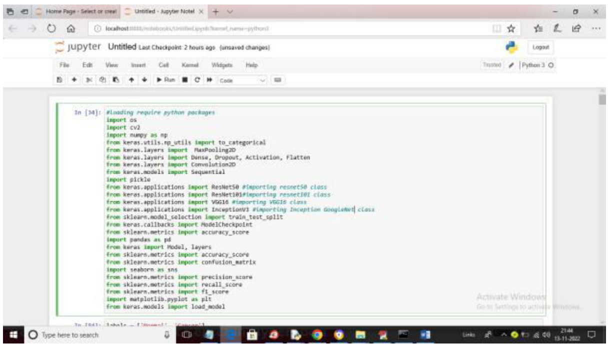
IV.System Model

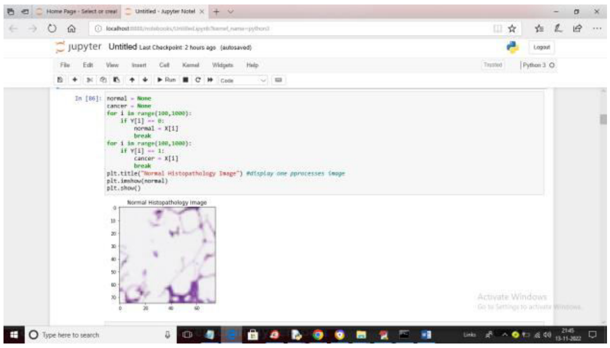
SYSTEM ARCHITECTURE

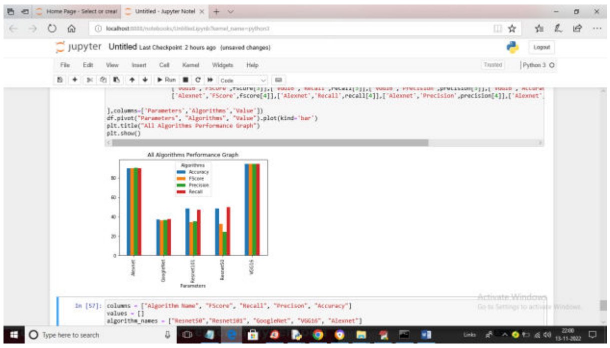


V.Results and Discussions









VI. CONCLUSION

The proposed system leveraging Cycle-Consistent Generative Adversarial Networks (Cycle GANs) for data augmentation in breast cancer diagnosis represents a substantial advancement in medical image analysis. By generating synthetic pathological images, this approach addresses the critical challenge of limited annotated data, which has traditionally hampered the development and accuracy of automated diagnostic models. The integration of Cycle GANs into the training process of convolutional neural networks (CNNs) offers a promising solution to enhance the performance and robustness of these models.

Cycle GANs provide a powerful tool for expanding the training dataset by creating realistic and diverse synthetic images that closely resemble actual pathological samples. This augmentation improves the model's ability to generalize across various tissue characteristics and abnormalities, leading to more accurate and reliable diagnostic predictions. The enhanced dataset not only mitigates the issues associated with data scarcity but also enriches the learning process by introducing variations that help the model distinguish between malignant and benign conditions more effectively.

Despite the notable advantages, the implementation of Cycle GANs is not without challenges. The computational resources required for training Cycle GANs can be substantial, and ensuring the quality and relevance of synthetic images is crucial to avoid introducing biases or artifacts. Rigorous validation is necessary to confirm that the synthetic images contribute positively to model performance and do not compromise diagnostic accuracy. Addressing these challenges is essential for maximizing the benefits of Cycle GANs in the diagnostic process.

VII. REFERENCES

1. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1455–1462. https://doi.org/10.1109/TBME.2015.2496264
2. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2223–2232). https://doi.org/10.1109/ICCV.2017.244
3. Baur, C., Albarqouni, S., & Navab, N. (2018). Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In *International MICCAI Brainlesion Workshop* (pp. 161–169). Springer, Cham. https://doi.org/10.1007/978-3-030-11726-9_16
4. Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, *58*, 101552. https://doi.org/10.1016/j.media.2019.101552