IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

BREAST CANCER DIAGNOSIS ON PATHOLOGICAL IMAGE DATA AUGMENTATION METHOD:CYCLE GAN

N.Noorjahan[1], C.Navya[2]

[1]M. Tech Student, Department of Computer science and engineering, Golden valley integrated campus, Kadiri Road, Angallu Post, Madanapalli, Chittoor, Andhra Pradesh 517326
[2]Assistant Professor, Department of Computer science and engineering, Golden valley integrated campus, Kadiri Road, Angallu Post, Madanapalli, Chittoor, Andhra Pradesh 517326

Abstract Breast cancer is one of the most common cancers affecting women worldwide, and early and accurate diagnosis is essential for improving patient survival. Histopathological image analysis plays an important role in breast cancer diagnosis, but the development of deep learning models is often limited by insufficient, imbalanced, and diverse pathological image datasets. This project, **"Breast Cancer Diagnosis on Pathological Images Using Data Augmentation Method: CycleGAN,"** proposes the use of **Cycle-Consistent Generative Adversarial Networks (CycleGAN)** to generate realistic synthetic pathological images for effective data augmentation. CycleGAN can learn the visual characteristics of breast tissue images and generate additional samples without requiring paired images. The generated images are combined with the original dataset to increase the diversity and quantity of training data. A deep learning-based classification model can then be trained on the augmented dataset to distinguish between benign and malignant breast tissue patterns. The proposed approach aims to reduce overfitting, improve model generalization, and enhance diagnostic classification performance. By incorporating synthetic pathological images into the training process, the system can provide a more robust and efficient computer-aided diagnosis approach. This methodology demonstrates the potential of generative deep learning techniques for addressing limited medical imaging datasets and supporting automated breast cancer diagnosis.

1 INTRODUCTION

Breast cancer is one of the most frequently diagnosed cancers among women and remains a major global health concern. Early detection and accurate diagnosis are important for selecting appropriate treatment and improving patient outcomes. Histopathological examination of breast tissue is an important diagnostic method in which tissue samples are examined under a microscope to identify abnormal cellular structures. However, manual analysis of pathological images is time-consuming and may be affected by variations in tissue appearance, staining techniques, image quality, and the experience of pathologists.

Recent developments in Artificial Intelligence (AI) and Deep Learning (DL) have provided effective methods for analyzing medical images automatically. Convolutional Neural Networks (CNNs) can learn complex visual patterns from pathological images and classify tissue into different categories. However, deep learning models generally require large and diverse datasets for effective training. In breast cancer pathological image analysis, obtaining sufficiently large and well-balanced datasets can be difficult because medical image collection and expert annotation require significant time and resources.

Data augmentation is commonly used to increase the size and diversity of training datasets. Traditional augmentation techniques such as rotation, flipping, cropping, scaling, and brightness adjustment can create additional samples, but they may not generate sufficiently realistic variations of pathological tissue structures. To overcome this limitation, this project proposes the use of Cycle-Consistent Generative Adversarial Network (CycleGAN) for pathological image augmentation.

CycleGAN is a generative deep learning model capable of learning image-to-image transformations without requiring paired training images. In the proposed system, CycleGAN is used to generate realistic synthetic breast pathological images that preserve important tissue characteristics while introducing additional visual variations. These generated images are combined with original pathological images to create an expanded dataset for training a breast cancer classification model.

The proposed approach aims to improve the diversity of the training data, reduce overfitting, and enhance the classification performance of breast cancer diagnosis systems. By combining CycleGAN-based data augmentation with deep learning-based image classification, the system provides an effective computer-aided approach for analyzing pathological images and supporting more reliable breast cancer diagnosis.

2. EXISTING SYSTEM

The existing breast cancer diagnosis system primarily depends on the manual examination of histopathological images by trained pathologists. In recent years, traditional machine learning and deep learning techniques have also been introduced to automatically classify pathological images as benign or malignant. These approaches generally use existing datasets to train classification models such as CNNs and other image-processing algorithms. Before training, conventional data augmentation techniques such as rotation, flipping, cropping, zooming, and brightness adjustment are commonly applied to increase the number of training images. Although these techniques improve dataset size to some extent, they mainly produce modified versions of existing images and may not provide sufficient diversity. In addition, pathological datasets are often limited, imbalanced, and affected by variations in staining, tissue structures, and image quality. When deep learning models are trained on small datasets, they may suffer from overfitting and poor generalization to unseen pathological images. Manual diagnosis can also be time-consuming and may involve inter-observer variation between pathologists. Therefore, conventional approaches may not fully utilize the potential of generative models for producing realistic pathological images. The absence of advanced synthetic image generation methods can limit the performance and robustness of automated breast cancer classification systems.

2.1 Disadvantages

- Limited availability of large and diverse pathological image datasets.
- Conventional augmentation produces limited variations of existing images.
- Class imbalance can negatively affect classification performance.
- Deep learning models may suffer from **overfitting** when trained on small datasets.
- Manual pathological examination is time-consuming.
- Diagnosis may vary depending on the experience of the pathologist.
- Variations in staining and image quality can affect model performance.
- Traditional methods may have poor generalization to unseen images.
- Generating realistic pathological variations using basic augmentation is difficult.

- Existing systems may not effectively utilize **GAN-based synthetic image generation** techniques.

3. PROPOSED SYSTEM:

The proposed system, **"Breast Cancer Diagnosis on Pathological Images Using Data Augmentation Method: CycleGAN,"** uses a Cycle-Consistent Generative Adversarial Network (CycleGAN) to improve the availability and diversity of breast cancer pathological image datasets. The system first collects and preprocesses pathological breast tissue images by resizing, normalization, and suitable image-quality processing. CycleGAN is then trained to learn the visual characteristics and variations present in the pathological images. Unlike traditional augmentation techniques, CycleGAN can generate new synthetic images with realistic tissue patterns without requiring paired images. These synthetic images are combined with the original pathological dataset to create an expanded and diverse training dataset. The augmented dataset is subsequently provided to a deep learning classification model for identifying pathological images into relevant breast cancer categories, such as benign and malignant. The model learns important cellular and tissue-level features from both original and generated images. The proposed approach aims to reduce overfitting and improve the generalization capability of the classification model. Performance can be evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. By integrating generative AI with pathological image analysis, the proposed system provides an effective computer-aided approach that can support pathologists in breast cancer diagnosis. The system is intended as a **diagnostic support tool**, rather than a replacement for clinical evaluation.

3.1 Advantages

- Generates realistic synthetic pathological images using **CycleGAN**.
- Increases the size of limited breast cancer image datasets.
- Improves dataset diversity and variability.
- Reduces the risk of **overfitting** during model training.
- Helps address class imbalance in pathological datasets.
- Produces image variations beyond basic rotation and flipping.
- Improves the generalization capability of the classification model.
- Supports automated identification of benign and malignant tissue patterns.
- Reduces dependence on manually collected additional images.
- Can improve classification performance when high-quality synthetic images are used appropriately.
- Provides quantitative evaluation using accuracy, precision, recall, and F1-score.
- Supports computer-aided breast cancer diagnosis and research.

4. SYSTEM ANALYSIS:

The proposed Breast Cancer Diagnosis on Pathological Images Using Data Augmentation Method: CycleGAN system is designed to overcome the limitations of conventional breast cancer pathological image analysis and traditional data augmentation techniques. The system initially collects pathological breast tissue images and performs preprocessing operations such as image resizing, normalization, quality checking, and class organization. The preprocessed images are then provided to the Cycle-Consistent Generative Adversarial Network (CycleGAN), which learns the visual characteristics and tissue patterns present in the dataset and generates realistic synthetic pathological images. These generated images are combined with the original images to create a larger and more diverse dataset. The augmented dataset is then used to train a deep learning-based classification model for identifying pathological images into categories such as benign and malignant.

The classification model learns important cellular and tissue-level features from the images and predicts the corresponding class for unseen samples. The performance of the proposed system can be evaluated using accuracy, precision, recall, F1-score, and confusion matrix. The overall system aims to improve dataset diversity, reduce overfitting, enhance classification performance, and provide a reliable computer-aided support system for breast cancer diagnosis. The proposed approach can be particularly useful when the availability of large, balanced, and diverse pathological image datasets is limited.

5. SYSTEM REQUIREMENTS:

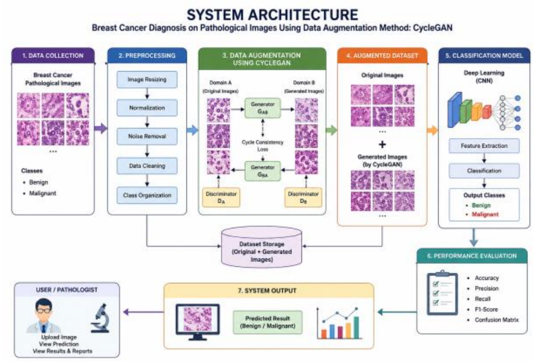
Hardware Requirements

- Processor: Intel Core i5 or equivalent
- RAM: Minimum 8 GB, recommended 16 GB
- Hard Disk: Minimum 500 GB
- GPU: NVIDIA GPU with minimum 4 GB VRAM (recommended)
- Monitor: 15.6-inch or higher
- Keyboard and Mouse
- Internet connection for dataset and software installation

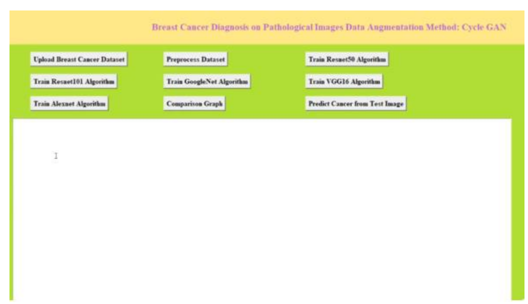
Software Requirements

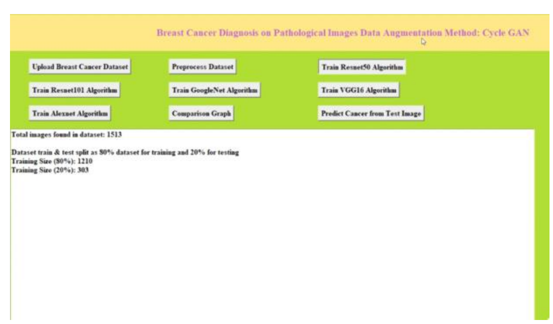
- Operating System: Windows 10/11 or Linux
- Programming Language: Python 3.10+
- Development Environment: Jupyter Notebook / VS Code
- Framework: TensorFlow / Keras
- Deep Learning Model: CycleGAN
- Image Processing: OpenCV, Pillow
- Data Processing: NumPy, Pandas
- Machine Learning: Scikit-learn
- Visualization: Matplotlib
- Database: MySQL 8.0 (if required for storing user and system information)
- Web Framework: Flask (if a web-based interface is implemented)

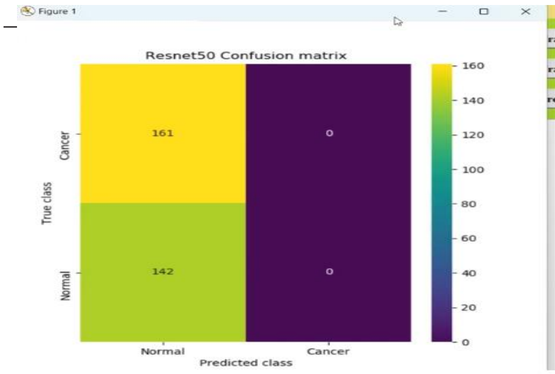
System Architecture:

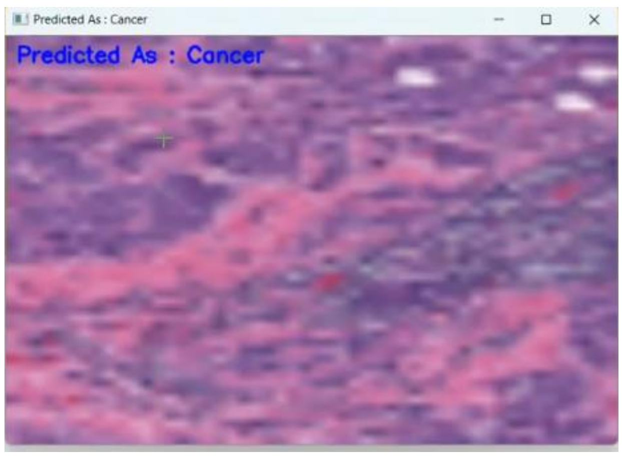


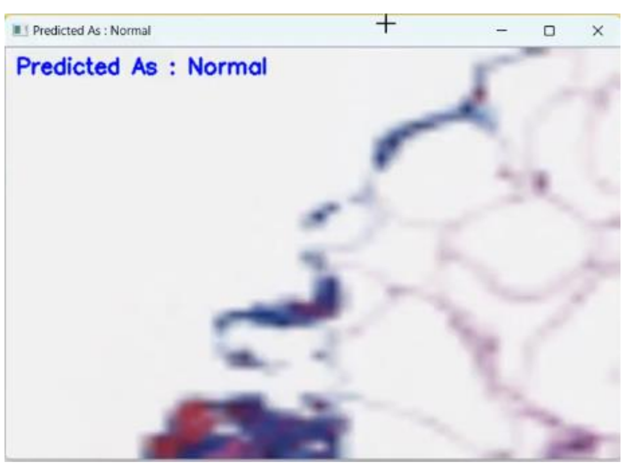
6. RESULTS AND ANALYSIS











7. CONCLUSIONS

The proposed Breast Cancer Diagnosis on Pathological Images Using Data Augmentation Method: CycleGAN provides an effective approach for improving automated breast cancer image classification. The system addresses the limitations of small, imbalanced, and less diverse pathological image datasets by using CycleGAN to generate realistic synthetic pathological images. These generated images are combined with original images to create an augmented dataset, which helps the deep learning classification model learn a wider range of tissue patterns and characteristics. The proposed approach can reduce overfitting and improve the model's ability to generalize to unseen pathological images. The system performs preprocessing, synthetic image generation, feature learning, classification, and performance evaluation in a systematic manner. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix can be used to measure the effectiveness of the classification model. Overall, CycleGAN-based data augmentation has strong potential to improve computer-aided analysis of breast pathological images and provide useful diagnostic support to medical professionals. However, the system should be considered a supporting tool rather than a replacement for expert pathological examination and clinical diagnosis.

REFERENCES

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems (NeurIPS)*, 27.

2. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2223–2232.

3. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisteds Intervention (MICCAI)*, 234–241.

4. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42, 60–88.

5. 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.

6. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., Marami, B., Prasanna, P., Sethi, A., Udupa, J. K., Depeursinge, A., & Carneiro, G. (2019). BACH: Grand Challenge on Breast Cancer Histology Images. *Medical Image Analysis*, 56, 122–139.

7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521, 436–444.

8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25.

9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

10. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708.

11. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the International Conference on Machine Learning (ICML)*, 6105–6114.

12. Shorten, C., &Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, 60.

13. Tellez, D., Balkenhol, M., Otte-Höller, I., van de Loo, R., Vogels, R., Bult, P., Wauters, C., Vreuls, W., Mol, S., van der Laak, J., &Ciompi, F. (2018). Whole-Slide Mitosis Detection in Breast Cancer Histology Images. *Medical Image Analysis*, 47, 109–119.

14. Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., &Litjens, G. (2020). Automated Deep-Learning System for Gleason Grading of Prostate Cancer Using Biopsies: A Diagnostic Study. *The Lancet Oncology*, 21(2), 233–241.

15. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A Guide to Deep Learning in Healthcare. *Nature Medicine*, 25, 24–29.