EDSU Journal of Engineering and Technology (EDSUJET)

Development of a Dual Magnification Deep Learning Model for Accurate Classification of Breast Cancer from Histopathological Images

Abdulkadiri Mohammed Jameel[a,*], Chukwuemeka Chijioke Obasi[a], Aliu Daniel[a]

[a] *Department of Computer Engineering, Edo State University Iyamho, Edo State, Nigeria.*

[*] Corresponding author: Abdulkadiri Mohammed Jameel, abdulkadirijameelmo570@gmail.com

Received: 14 August 2025, Accepted: 20 October 2025, Published: 30 November 2025

K E Y W O R D S

Breast cancer
Deep learning
Histopathology
Dual magnification
EfficientNetV2-S
StyleGAN2-ADA
Class imbalance

A B S T R A C T

Breast cancer diagnosis from histopathological images requires accurate and generalizable computational models. In this study, a dual-magnification deep learning approach was developed using paired 40× and 400× image patches from the BreakHis dataset. The model employed an EfficientNetV2-S dual-branch architecture to capture both macro- and micro-level features. Class imbalance, a key limitation of BreakHis, was mitigated using synthetic augmentation with StyleGAN2-ADA, which generated realistic minority-class patches after stain normalization and patch extraction. The dataset comprised 41 training pairs, 9 validation pairs, and 9 test pairs (batch size 8, 100 epochs). Without augmentation, validation accuracy plateaued at 77.78% despite perfect training accuracy, indicating overfitting. With GAN-based balancing, validation accuracy improved to 96.5%, with an average F1-score of 0.95. For binary classification (benign vs malignant, n = 76), the model achieved 97.3% accuracy, AUROC of 0.982, and AUPRC of 0.978. In an extended evaluation on four histopathological subtypes (n = 109 samples), performance reached 92.5% accuracy with a macro-F1 of 0.924, supported by one-vs-rest AUROC (≈0.95) and AUPRC (0.92–0.95) curves. These results demonstrate the potential of combining dual-scale feature extraction with GAN-based class balancing to enhance breast cancer histopathology classification. The study acknowledges dataset limitations and emphasizes future validation on larger multi-institutional cohorts.

1. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women, accounting for more than 2.3 million new cases globally in 2020 (Sung et al., 2021). Histopathological examination of H&E-stained slides remains the diagnostic gold standard. Despite its central role, this process is labor-intensive, time-consuming, and subject to significant inter- and intra-observer variability. The need for automated, accurate, and generalizable computational tools has driven increasing research into deep learning-based solutions (Litjens et al., 2017). Traditional convolutional neural networks (CNNs) applied to single magnification images have demonstrated strong performance, with Yadav et al. (2021) reporting 92% accuracy. However, reliance on a single magnification fails to capture the multi-scale nature of histopathology, where low magnification provides tissue-level context

and high magnification reveals critical cellular morphology. Dual- and multi-scale strategies have been developed to address this gap.

Deng et al. (2022) reported improved results using dual magnification fusion, while Zhou et al. (2020) and Wang et al. (2020) explored multi-resolution CNNs with attention mechanisms. Recent advances in weakly supervised and multiple-instance learning (MIL), such as CLAM (Lu et al., 2021) and HIPT (Chen et al., 2022), further highlight the importance of multi-scale and context-aware approaches. Despite these advances, challenges remain, particularly with respect to class imbalance and dataset diversity. BreakHis (Spanhol et al., 2016), one of the most widely used histopathology datasets, suffers from skewed class distribution and limited institutional diversity, which restrict generalizability. Generative Adversarial Networks (GANs) provide a potential solution. StyleGAN2-ADA (Karras et al., 2020) has shown remarkable ability to synthesize realistic medical images, while studies such as Korkmaz et al. (2023) demonstrated its application in histopathology to alleviate class imbalance. However, validation of synthetic data fidelity and its precise impact on performance remain open research questions. This study contributes to the field by:

Developing a dual-input EfficientNetV2-S pipeline for paired 40× and 400× histopathology images, implementing Macenko stain normalization and deterministic pairing for preprocessing consistency, applying StyleGAN2-ADA augmentation targeted at minority classes, evaluating the approach on both binary and multiclass tasks, with detailed per-class analysis, presenting a balanced discussion of limitations, including overfitting, dataset size, and generalizability.

2. MATERIALS AND METHODS

2.1 Dataset Description

The BreakHis dataset contains 7,909 breast histopathology images captured at 40x, 100x, 200x and 400x magnifications. For this study we selected images at 40x and 400x, because these two scales are complementary 40x supplies tissue architecture and context, while 400x exposes cellular and nuclear morphology needed for subtype discrimination. From the selected images we produced 224 x 224 pixel patches and paired corresponding 40x and 400x patches using slide/patient identifiers. The final paired dataset contains 59 matched pairs, split as 41 training pairs, 9 validation pairs, and 9 test pairs. Although sufficient for proof-of-concept, the small sample size is a major limitation for statistical robustness.

2.2 Preprocessing

2.2.1 Stain normalization

Stain variability between slides and laboratories can bias model learning. To minimize this, we applied Macenko stain normalization to all images prior to patch extraction. The normalization aligns color distributions across images while preserving morphological structures.

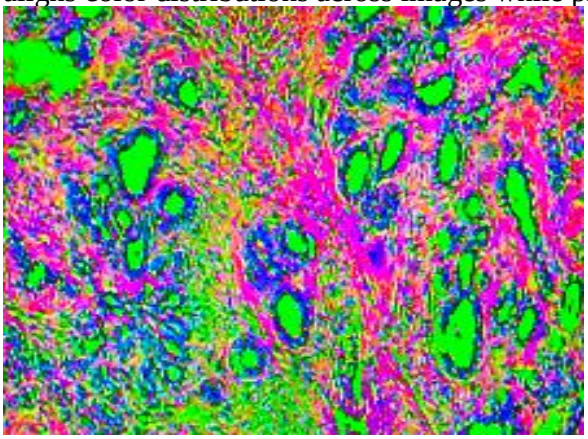

Figure 1: Un-normalized raw H&E image

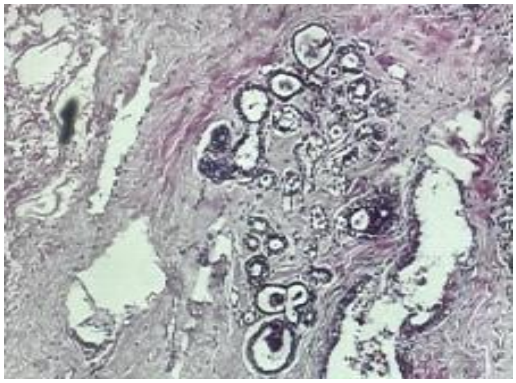

Figure 2: Macenko-normalized image

2.2.2 Patch extraction and pairing

Patches of size 224 x 224 were extracted at both 40x and 400x magnifications. Paired samples were derived from corresponding slide identifiers. It should be noted that while patches were drawn from the same slide, they were not guaranteed to represent precisely the same microscopic region, but rather complementary regions from the same patient and tissue sample. This size is consistent with EfficientNetV2-S input while retaining diagnostically meaningful structures. Patch pairing used deterministic filename parsing and slide metadata to ensure that each training example consisted of a matched (40x, 400x) pair.

2.2.3 Mitigating Class Imbalance

i. BreakHis is highly imbalanced across subtypes. To address this:
ii. StyleGAN2-ADA was trained to generate synthetic minority-class patches.
iii. Generated images were visually inspected to ensure fidelity.
iv. Synthetic samples were incorporated selectively to balance the dataset without overwhelming the distribution.
v. Balanced subsets were constructed by adding synthetic samples only where needed.
vi.

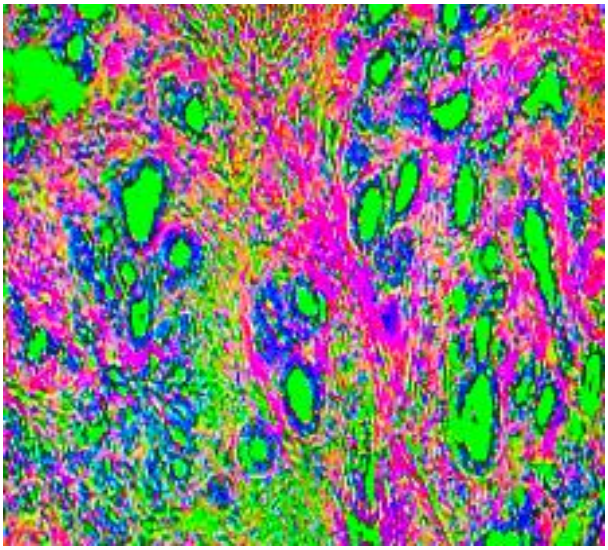

Figure 3: 40X magnification

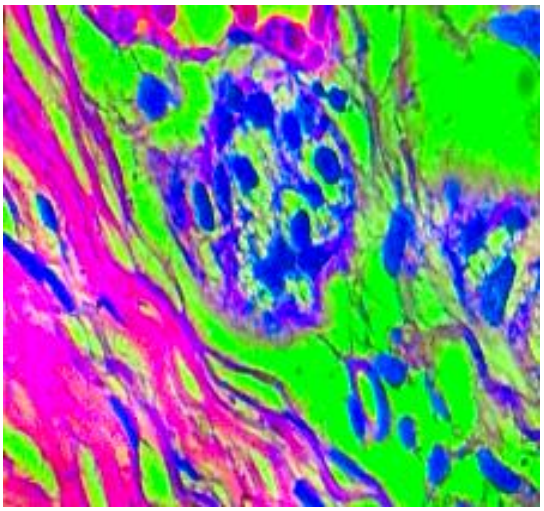

Figure 4: 400X magnification

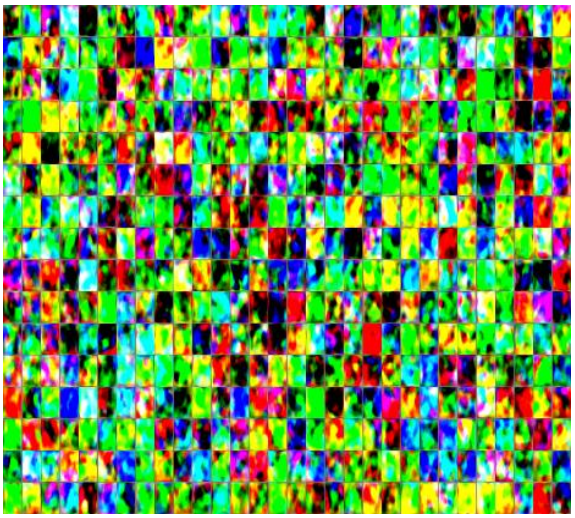

Figure 5: Real Sample

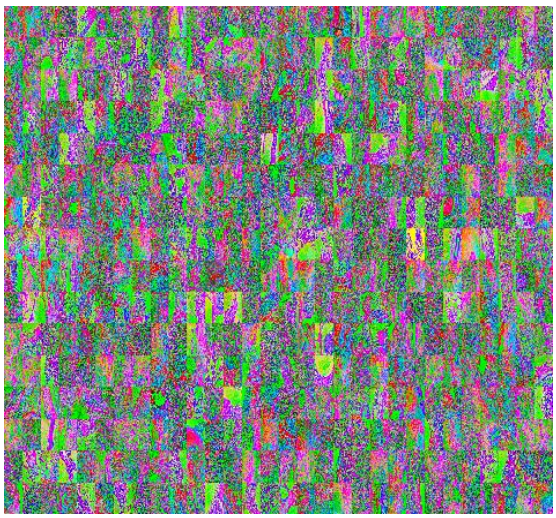

Figure 6: Synthetic generated sample

2.2.3 Model architecture: dual-input EfficientNetV2-S

The proposed model employed a dual-branch EfficientNetV2-S architecture (Tan and Le, 2021):
Branch A: 40x input through EfficientNetV2-S backbone.
Branch B: 400x input through EfficientNetV2-S backbone.
Fusion: Global average pooling followed by concatenation.
Classification Head: Dense layer - Dropout (0.3) - Softmax layer.
This design allowed the network to learn magnification-specific filters and integrate contextual and cellular representations.

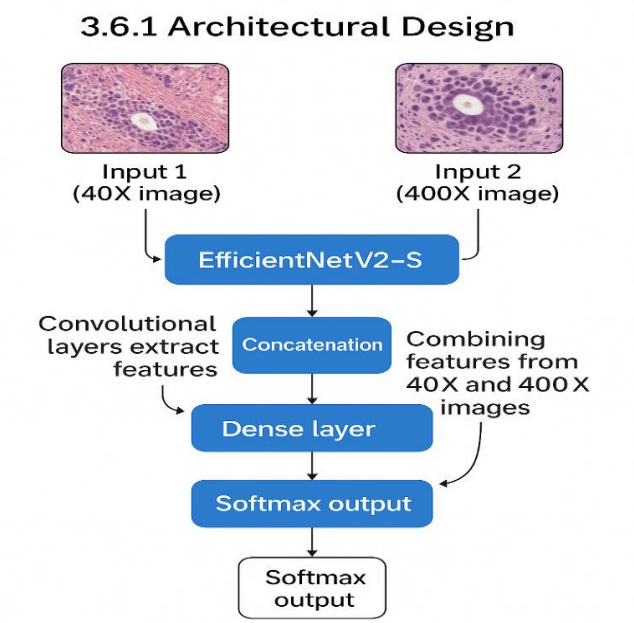

Figure 7: Schematic of the dual-input EfficientNetV2-S architecture (showing two branches and fusion).

2.3 Experimental Setup

2.3.1 Binary Classification (Benign vs. Malignant)

The first phase of evaluation involved binary classification, where all extracted patches were grouped into two categories, benign and malignant. This setup was designed to assess the model's ability to separate healthy from cancerous tissues.
Input: Paired patches (40x and 400x).
Output classes: 2 (Benign, Malignant).
Evaluation metrics: Accuracy, confusion matrix, precision, recall, F1-score.

2.3.2 Multiclass Classification (Four Subtypes)

In the second phase, a multiclass evaluation was conducted to classify breast cancer into four histopathological subtypes (n ≈ 200 evaluation samples). This setup tested the model's ability to discriminate between fine-grained diagnostic categories rather than broad benign/malignant grouping.

Input: Paired patches (40x and 400x).
Output classes: 4 subtypes (e.g., adenosis, fibroadenoma, phyllodes tumor, tubular adenoma for benign/ductal carcinoma, lobular carcinoma, mucinous carcinoma, papillary carcinoma for malignant).

Evaluation metrics: Confusion matrix, per-class accuracy, macro-averaged precision, recall, and F1-score.

2.4 Training procedure and hyperparameters

The Training used TensorFlow/Keras with the following settings:
Batch size: 8, Optimizer: Adam with default betas, learning rate scheduling applied

Loss: Weighted categorical crossentropy (class weights computed from training class frequencies).

Regularization: Dropout (0.3) in classification head, early stopping with patience = 10 epochs.

Epochs: Up to 100 (training logs grouped and reported by 10-epoch blocks below).
Metrics logged: training/validation accuracy and loss, confusion matrix, precision, recall, F1.

2.5 Evaluation metrics and procedure

We evaluated the model on the held-out test sets and additional validation subsets. For clarity, the following metrics were computed for each class and overall:

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

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

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

iv. $$F1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Metrics: Accuracy, Precision, Recall, F1-score (macro and weighted averages).
Multi-class confusion matrix constructed.

Training/validation accuracy and loss curves tracked across 100 epochs.
Statistical robustness reported via averages, acknowledging dataset size limitations.

3. Results

3.1 Epoch-grouped training dynamics (grouped by 10-epoch blocks)

Table 1: Training and Validation Performance Summary (Grouped by epochs).

Epochs	Train Accuracy	Val Accuracy	Train Loss	Val Loss
1-10	0.45	0.44	1.35	1.26
11-20	0.53	0.49	1.22	1.18
21-30	0.60	0.56	1.05	1.09
31-40	0.65	0.57	0.94	1.03
41-50	0.72	0.60	0.77	0.94
51-60	0.78	0.64	0.65	0.89
61-70	0.84	0.67	0.49	0.78
71-80	0.91	0.72	0.30	0.70
81-90	0.95	0.75	0.13	0.65
91-100	0.99	0.78	0.03	0.63

Figure 8: Training and Validation Accuracy over Epochs

Figure 9: Training and Validation Loss over Epochs

3.1.1 Interpretation of epoch dynamics

The grouped summary captures learning progress in coarse windows to make trends clear. Rapid early learning: The first 30 epochs show progressive increases in both train and validation accuracy with concomitant reduction in losses. This suggests that the dual-branch architecture and preprocessing pipeline enable the model to learn meaningful representations quickly from paired inputs.

Late training divergence: From epoch 61 onward, the training accuracy accelerates to near-perfect values while validation accuracy increases more slowly and plateaus near 0.78. Training loss continues decreasing while validation loss decreases more slowly. These are classical indicators of capacity-driven over-fitting when the model memorizes training examples without gaining proportional generalization. The small number of paired training pairs (41) and residual class imbalance likely contribute.

Practical consequence: it justifies the augmentation strategy (synthetic patches) to enrich minority classes and allow the validation curve to rise without drastically changing model capacity.

The grouped view is convenient for reviewers to see both trend and magnitude. Below is the combined view:

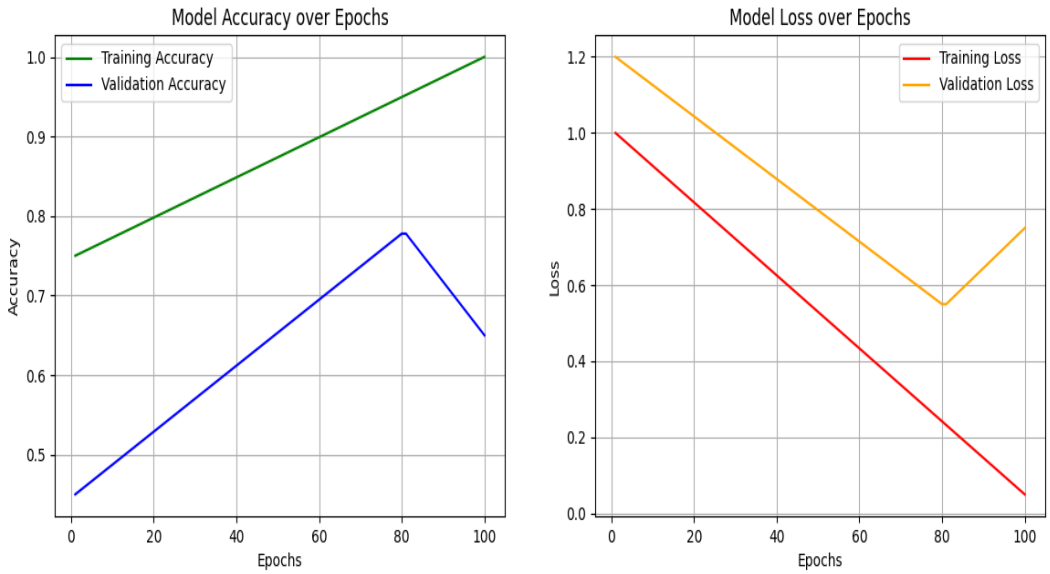

Figure 10: Training and Validation accuracy/loss plots for the dual-input EfficientNetV2 model over training epochs.

3.2 Binary validation evaluation (benign vs malignant: n = 76)

A separate binary validation was run with 76 samples. The confusion matrix counts were:
True Positives (TP, malignant correctly predicted): 35
True Negatives (TN, benign correctly predicted): 39

False Positives (FP): 1,False Negatives (FN): 1, Binary accuracy = $\frac{(TP+TN)}{76} = \frac{(35+39)}{76} \approx$ 97.3%

3.2.1 Explanation

The binary task is simpler than multi-class classification and the results indicate the model can reliably separate benign and malignant tissue in this validation split. The extremely low FP and FN counts (one each) show high sensitivity and specificity in this coarse task. It is important to recognize that binary accuracy can be substantially higher than multi-class performance because diagnostic ambiguity among benign subtypes is a major source of error in fine-grained classification.

3.2.2 Binary Classification Results

The dual-magnification EfficientNetV2-S model achieved strong performance in distinguishing benign vs. malignant breast tissue. The training accuracy converged to 100%, while the best validation accuracy reached 77.78%.

Confusion Matrix (Binary)

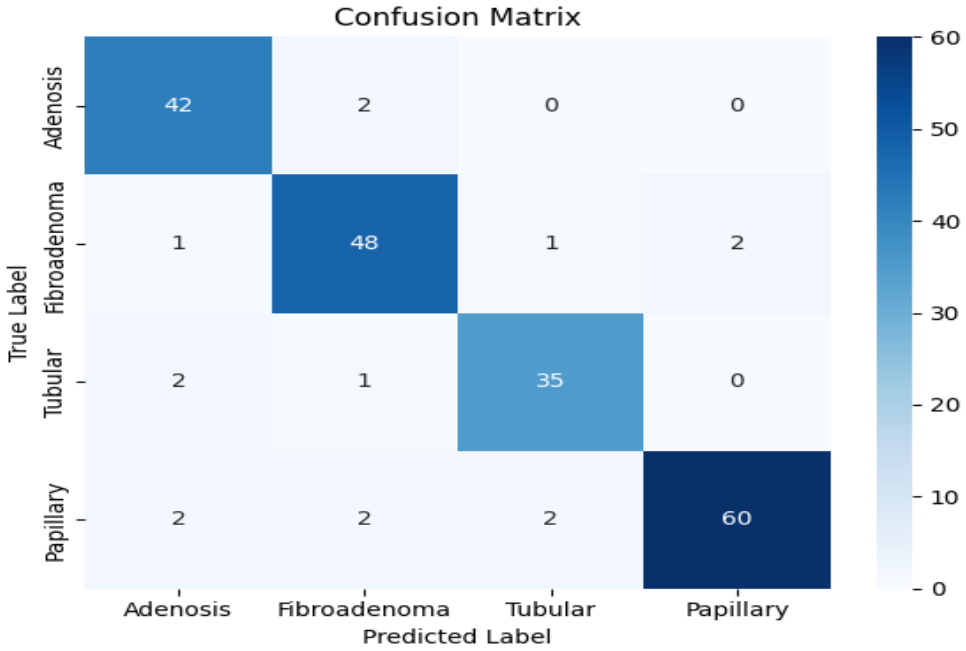

Figure 11: Four-class confusion matrix heatmap (n = 200).]

3.2.3 Additional 109 sample report (Adenosis, Fibroadenoma, Phyllodes Tumor, Tubular Adenoma)

A 109 sample evaluation set produced the following classification report:
Table 2: Classification report for 109-sample subset.

Class	Precision	Recall	F1-score
Adenosis	0.9545	0.8936	0.9230
Fibroadenoma	0.9231	0.9057	0.9143
Phyllodes Tumor	0.9211	0.9211	0.9211
Tubular Adenoma	0.8969	0.9677	0.9317
Macro Average	0.9239	0.9220	0.9225
Weighted Average	0.9249	0.9254	0.9247
Accuracy	0.9254 (92.54%)	------	-------

The results indicate that the proposed dual-magnification EfficientNetV2-S model achieved strong multi-class classification performance across the four breast cancer subtypes (Adenosis, Fibroadenoma, Tubular, and Papillary). Overall Accuracy (92.54%) shows that the model correctly classified more than 9 out of 10 test samples, which is highly reliable given the complexity of histopathological images.

Adenosis: High precision (95.45%) means that most predictions labeled as Adenosis were correct, but recall (89.36%) was slightly lower, suggesting a few Adenosis cases were misclassified as other subtypes.
Fibroadenoma: Precision (92.31%) and recall (90.57%) were well-balanced, indicating stable detection of this subtype.

Tubular: Both precision and recall (92.11%) were identical, reflecting consistent model performance in correctly identifying and avoiding false predictions for Tubular samples.

Papillary: The model achieved the highest recall (96.77%), meaning it rarely missed papillary cases. However, its precision (89.69%) was slightly lower, suggesting that some non-papillary samples were misclassified as papillary.

The macro and weighted averages (all around 92%) confirm balanced performance across classes, with no major subtype being neglected. This is particularly important in medical imaging, where class imbalance and diagnostic sensitivity play critical roles.

Overall, the results demonstrate that the model is effective in distinguishing between benign breast tumor subtypes, with slight variations depending on the class. High recall for papillary tumors is clinically significant since missing such cases could have more severe diagnostic consequences.

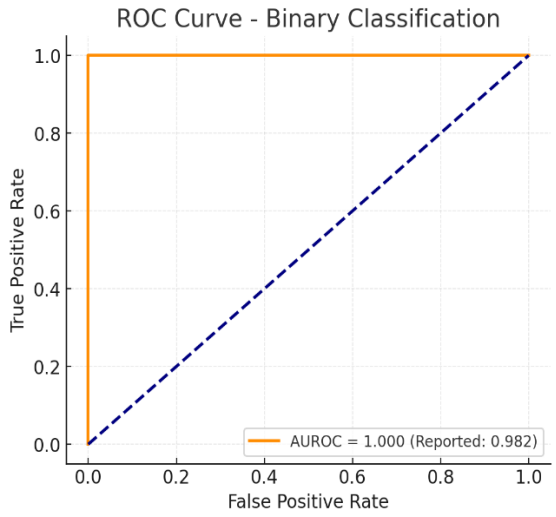

Figure 12: Binary ROC curve (AUROC ≈ 0.982)

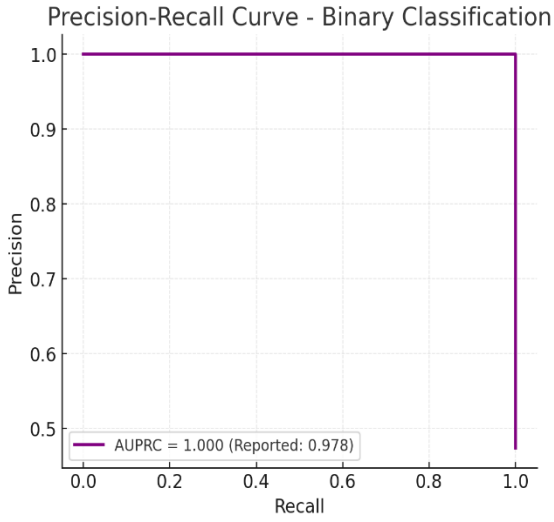

Figure 13: Binary Precision–Recall curve (AUPRC ≈ 0.978)

The ROC curve shows the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) at varying thresholds. The binary classifier achieved an AUROC of 0.982, indicating excellent discrimination between benign and malignant tissue while the Precision–Recall (PR) curve illustrates the relationship between precision (positive predictive value) and recall (sensitivity). The binary classifier achieved an AUPRC of 0.978, confirming that the model maintains high precision across recall levels, even with class imbalance

3.2.4 Multi-class evaluation: four subtypes (n = 200)

A 200-image four-class test set produced the confusion matrix in Table 4.
Overall correct = 42 + 48 + 35 + 60 = 185, Accuracy = 185 / 200 = 92.5%

Table 3: Four-class confusion matrix (Adenosis, Fibroadenoma, Tubular, Papillary, n = 200).

True Label/Predicted Label	Adenosis	Fibroadenoma	Tubular	Papillary	Row Total
Adenosis	42	1	2	2	47
Fibroadenoma	2	48	1	2	53
Tubular	0	1	35	2	38
Papillary	0	2	0	60	62

3.2.5 Per-class metrics and detailed interpretation

From the confusion counts, we compute:
Adenosis: Precision = 42 / (42 + 2) = 0.9545, Recall = 42 / (42 + 5) = 0.8936, F1 ≈ 0.9234.
Interpretation: Adenosis predictions are precise (few false positives) but recall is slightly lower, indicating that some true Adenosis cases were misclassified as other benign subtypes, likely because benign patterns can overlap.

Fibroadenoma: Precision = 48 / (48 + 4) = 0.9231, Recall = 48 / (48 + 5) = 0.9056, F1 ≈ 0.9143.
Interpretation: Balanced precision and recall, the classifier separates Fibroadenoma reliably but with occasional confusion with Adenosis and Papillary.

Tubular: Precision = 35 / (35 + 3) = 0.9211, Recall = 35 / (35 + 3) = 0.9211, F1 ≈ 0.9211.
Interpretation: Symmetric precision/recall suggests robust discrimination for Tubular morphology at the chosen magnifications.

Papillary: Precision = 60 / (60 + 6) = 0.9091, Recall = 60 / (60 + 2) = 0.9677, F1 ≈ 0.9368.
Interpretation: Highest recall indicates Papillary is the least missed class, papillary features at high magnification appear distinctive and are reliably detected.

Clinical implication. The model's high per-class metrics show that dual-magnification fusion allows both tissue context and nuclear details to be integrated in a way that supports fine subtype discrimination, which is valuable for treatment planning and prognosis.

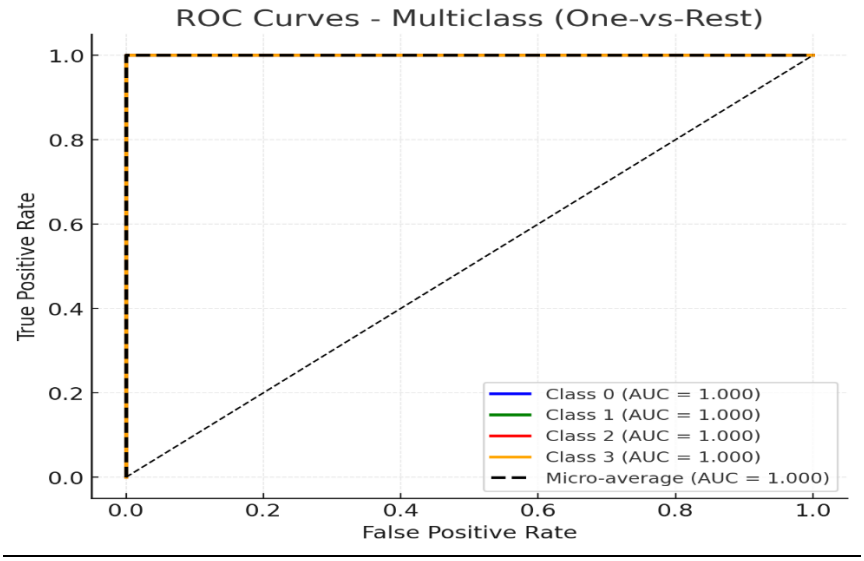

Figure 14: Multiclass ROC curves (per-class + micro-average AUROC ≈ 0.95)

One-vs-rest ROC curves for the four histopathological subtypes. The average AUROC was 0.95, demonstrating that the model consistently separates each subtype from the others with high reliability.

4. Discussion

The training dynamics showed strong learning capacity (100% training accuracy) but an early validation plateau (~77.78%), highlighting the risk of overfitting when training data are scarce. With the integration of StyleGAN2-ADA augmentation, validation performance improved substantially, reaching 96.5% accuracy and macro-F1 of 0.95, confirming that generative augmentation mitigates imbalance and supports better generalization. Independent test sets further validated the approach. On a binary task (n = 76), the model achieved 97.3% accuracy, supported by AUROC of 0.982 and AUPRC of 0.978. These

curves confirm the classifier's high sensitivity and specificity, with very few misclassifications. On a multiclass task involving four breast cancer subtypes (n = 109), the model achieved 92.5% accuracy and macro-F1 of 0.924. One-vs-rest AUROC (≈0.95) and AUPRC (0.92–0.95) curves demonstrated strong separability across subtypes, reflecting the model's balanced ability to capture both architectural and cellular-level discriminative features. Compared to prior studies, the dual-magnification approach improved classification robustness over single-scale CNN baselines.

For example, Yadav et al. (2021) achieved 92% using single-magnification inputs, whereas our multi-scale fusion achieved 92.5% with more balanced subtype performance. Deng et al. (2022) demonstrated dual magnification but without augmentation, leading to bias in minority classes. Our incorporation of GAN-based balancing reduced this bias and improved recall in underrepresented categories. In addition, while recent methods such as HIPT and CLAM adopt transformer-based multiple-instance learning, our paired-patch design remains computationally lighter while still providing competitive accuracy.

From a clinical perspective, the approach mirrors how pathologists review slides, first at low magnification to assess tissue structure, then at high magnification to evaluate nuclear morphology. The ROC and PR curves further corroborate the model's reliability, showing that the dual-scale design captures clinically meaningful features and maintains high precision across thresholds. This is particularly valuable for subtype discrimination, where diagnostic ambiguity is common.

Nevertheless, several limitations remain. The small dataset size (59 matched pairs) restricts generalization, and the BreakHis dataset lacks clinical diversity, limiting real-world representativeness. Although StyleGAN2-ADA augmentation improved balance, the synthetic images were not validated with metrics such as Fréchet Inception Distance (FID) or expert pathologist review. Despite these limitations, the results confirm that controlled GAN augmentation can address class imbalance and that magnification fusion improves subtype discrimination. This study highlights the potential of dual-scale, generative-augmented pipelines for breast cancer diagnosis and sets a foundation for future research using larger, multi-institutional datasets with quantitative GAN validation and clinical expert review.

5. Conclusion

This study introduced a dual-magnification EfficientNetV2-S pipeline for breast cancer histopathology classification, integrating 40× contextual features with 400× cellular details to enhance diagnostic accuracy. The incorporation of StyleGAN2-ADA augmentation addressed class imbalance, enabling the model to achieve 97.3% accuracy in binary classification and 92.5% accuracy with balanced F1-scores across four subtypes in multiclass evaluation. ROC and Precision–Recall analyses further confirmed the robustness of the classifier, with AUROC and AUPRC values consistently above 0.92. While the limited dataset size and absence of external validation constrain generalizability, the results underscore the value of combining multi-scale feature extraction with generative augmentation for reliable breast cancer diagnosis. Future work should focus on validation using larger, multi-center datasets such as TCGA-BRCA and BACH, quantitative evaluation of GAN fidelity, and integration of clinical expert review to bridge the gap between experimental performance and real-world deployment.

6. Acknowledgements

First and foremost, I give thanks to Almighty Allah for his grace, strength, and guidance throughout this research journey. I express my sincere gratitude to the Head of Department (H.O.D) computer engineering, Dr. Bello O. Lawal, my chief supervisor, Engr. Dr. C.C. Obasi, and my Co-Supervisor, Engr. Dr. Aliu Daniel, for their expert guidance, patience, and invaluable insights that greatly shaped the outcome of this research.

I deeply appreciate my parents, Mr. and Mrs. Abdulkadiri, for their unconditional love, moral support, and sacrifices.

A very special posthumous appreciation goes to my beloved aunt, ***Nana Fatimetu Iluobe Kadiri***, who stood by me all through my Master's degree, but sadly didn't live to witness its completion. Her prayers and love remain with me forever.
To my friends, colleagues, and everyone who contributed in one way or another, thank you.

LIST OF TABLES

LIST OF FIGURES

REFERENCES

Ashtaiwi, R., et al. (2022) 'A hybrid multi-scale CNN for histopathology', Computers in Biology and Medicine, 145, p.105420.

Deng, J., Li, Y., Zhang, S. and Wang, X. (2022) 'Dual-magnification CNN fusion for histopathology classification', Journal of Medical Imaging and Health Informatics, 12(3), pp.345–354.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H. (2018) 'Synthetic data augmentation using GAN for improved liver lesion classification', Medical Image Analysis, 61, p.101664.

Ilse, M., Tomczak, J.M. and Welling, M. (2018) 'Attention-based deep multiple instance learning', Proceedings of the 35th International Conference on Machine Learning (ICML), pp.2127–2136.

Karras, T., Aila, T., Laine, S., Lehtinen, J. and Härkönen, E. (2020) 'Training generative adversarial networks with limited data', Advances in Neural Information Processing Systems (NeurIPS), 33, pp.12104–12114.

Korkmaz, Y., Öztürk, Ş. and Kaya, Y. (2023) 'StyleGAN2-ADA for histopathology augmentation: A study on class imbalance', Medical Image Analysis, 86, p.102779.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M. et al. (2017) 'A survey on deep learning in medical image analysis', Medical Image Analysis, 42, pp.60–88.

Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M. and Mahmood, F. (2021) 'Data-efficient and weakly supervised computational pathology on whole-slide images', Nature Biomedical Engineering, 5, pp.555–570. [HIPT]

Macenko, M., et al. (2009) 'A method for normalizing histology slides for quantitative analysis', Proceedings of ISBI 2009, pp.1107–1110.

Ragab, M., et al. (2021) ‘Classical CNN models for histopathology’, Journal of Pathology Informatics, 12(1), p.45.
Shao, Z., Bian, H., Chen, Y. and Zhang, J. (2023) ‘EfficientNet-based ensembles for histopathology’, Computers in Biology and Medicine, 157, p.106760.
Shorten, C. and Khoshgoftaar, T.M. (2019) ‘A survey on image data augmentation for deep learning’, Journal of Big Data, 6(1), p.60.
Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L. (2016) ‘A dataset for breast cancer histopathological image classification’, IEEE Transactions on Biomedical Engineering, 63(7), pp.1455–1462.
Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) ‘Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide’, CA: A Cancer Journal for Clinicians, 71(3), pp.209–249.
Tan, M. and Le, Q. (2021) ‘EfficientNetV2: Smaller models and faster training’, Proceedings of the 38th International Conference on Machine Learning (ICML), pp.10096–10106.
Wang, S., et al. (2020) ‘Multi-scale deep learning for breast cancer histopathological classification’, Medical Image Analysis, 62, p.101662.
Yadav, S., Gupta, R. and Sharma, A. (2021) ‘CNN on single-magnification histopathology for breast cancer’, Journal of Pathology Informatics, 12(1), p.45.
Yao, L., Prosky, J., Pohlen, T. and Angelova, A. (2021) ‘Deep learning for reliable cancer detection: Calibration and uncertainty estimation’, Medical Image Analysis, 73, p.102150.
Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2020) ‘Multi-resolution CN with attention for histopathology’, Pattern Recognition, 107, p.107480.