Magnification Independent Multi-Classification of Breast Cancer in Histopathology Images Using Deep Learning

Jason Nishio
BASIS Independent Silicon Valley
San Jose, USA
jasonnishoxd@gmail.com

Nelson Nishio
BASIS Independent Silicon Valley
San Jose, USA
nelsonknisho@gmail.com

Alexander Nishio
Department of Computer Science
Purdue University
West Lafayette, USA
anishio@purdue.edu

***Abstract*—Breast cancer is one of the most common and leading causes of female mortality in the world. The disease can be treated quickly and successfully if the tumor type is diagnosed at an early stage. Histopathology provides superior breast cancer analysis. However, due to tumor cell heterogeneity, manually identifying and classifying breast tumors in histopathological images is highly labor-intensive for pathologists and can sometimes be inaccurate. Hence, there is a need to develop an automated detection framework using deep learning. This study leverages the pre-trained ResNet50 and VGG19 models for detecting breast cancer in histopathological images from the BreakHis dataset. Unlike previous research that often separated images by magnification for classification, this study bridges this gap by constructing models for magnification-independent multi-tumor-type classification. The key success of this study is the proposed Z-score color normalization in RGB channels during image preprocessing, which reduces variations in color, brightness, and contrast. The pre-trained ResNet-50 and VGG19 models achieved 99.03% and 96.59% accuracies in binary classification, and 94.73% and 92.50% accuracies in multi-classification, respectively. These results surpass the performance of existing methods in magnification-independent multi-classification for breast cancer detection.**

Keywords—breast cancer, histopathological images, mitosis, deep learning, ResNet50, VGG19, BreakHis

I. Introduction

Breast cancer (BC) is the second leading cause of female mortality globally, surpassing lung cancer in 2022 [1]. Although breast cancer can be treated through surgical removal, teletherapy, and radiation therapy, a definitive cure remains elusive with current technology. Early detection of cancer is crucial in preventing breast cancer from advancing to a severe stage, enabling doctors to identify and remove early-stage cancer cells promptly, thereby significantly decreasing the mortality rate associated with breast cancer. Mammograms, which use low-dose radiation, are typically used for breast cancer screening in asymptomatic women. However, manual analysis of image texture and morphology presents challenges due to noise and resolution limitations, compromising diagnostic accuracy [2]. Advanced medical techniques such as histopathology provide superior breast cancer analysis. When abnormalities are detected on mammograms, biopsies are conducted, and breast histopathology images – stained with Hematoxylin and Eosin (H&E) – are examined by pathologists through a microscope at high magnification, usually 40× or higher. Manual mitotic counts, which serve as indicators of cell proliferation rates and cancer aggressiveness, are labor-intensive and challenging for pathologists, as they require identifying and tallying mitotic figures. The accuracy of these counts can be influenced by factors such as the pathologist's expertise, sampling biases, and fatigue. Automated mitosis detection systems can mitigate these issues, providing more reliable results. With the impressive track record of deep learning in visual recognition tasks, there is significant interest in developing tools to assist pathologists and enhance the precision of mitotic detection in histopathological images.

Automated mitosis detection faces three primary challenges based on various normal and abnormal mitosis patterns identified in [3]. First, mitosis typically progresses through four major phases—prophase, metaphase, anaphase, and telophase—each exhibiting significant variations in cell shape and size due to the intricate structure of human body cells. This complexity hinders mitosis detection using computational algorithms, which normally rely on learning patterns in shape, size, texture, and other features. Second, histopathology images contain numerous cellular structures that resemble mitosis but are not true mitotic events. These mimics can lead to an increase in false positives during cell detection. Third, variations in staining colors, cell density, magnification scales and image quality across histology slides—due to differences in staining processes, scanners and microscopes—can impact the accuracy of tumor classification by deep learning models. Therefore, there is an urgent need to develop automated and effective BC diagnosis and prognostic techniques to address the rising number of untreated cases and alleviate the workload on healthcare providers.

II. Related Work

Several methods for BC detection have been proposed in previous research. Among them, transfer learning is one of the most popular deep learning methods used in BC detection research [4-6]. Transfer learning involves utilizing pre-trained neural networks that have been trained on large datasets as the starting point for a new training task, requiring less computing effort and expense. Various deep learning networks, such as VGG [4][5][7], ResNet [4][8], and DenseNet [6][9], have been developed and are widely used for the analysis of medical images.

Most research efforts on breast cancer using histopathological images have focused on binary classification. In 2023, Kanavos and Mylonas [4] explored the use of two popular pre-trained networks, VGG and ResNet. They retained the main structure of the pre-trained networks, excluding the last fully connected layer. The weights of these networks were frozen during training, assuming they have been optimized for a large amount of data from previous usage. A new fully connected layer was then added. The two optimizers, SGD and Adam, along with the hyperparameters, batch size, and epochs, were evaluated. They reported that their proposed ResNet architecture, optimized with the SGD optimizer, achieved the best results among the tested models, with an accuracy of 94.1% and an F1-score of 94.4% in breast cancer detection. Using the BreakHis dataset, Fauzan et al. [6] compared three pre-trained computer vision models—DenseNet, RegNet, and BiT—in predicting malignant or benign tumor tissue from breast histopathology images. They suggested that the BiT model was more suitable for deployment in a real-world setting, as it could predict more malignant cases correctly than the other two models, with a sensitivity score of 90.79%, although the DenseNet model achieved the highest average precision score of 97.4%.

While the studies above primarily focused on binary classification, a few approaches have been proposed for multi-class classification of eight breast tumor types from histopathological images [7-9]. Sharma et al. [7] and Agaba et al. [8] both separated images into different magnification levels for multi-classification to reduce variability in microscopy images. They employed handcrafted feature extraction techniques, such as Hu moments for cell shapes, Haralick textures for cell surface and appearance, and color histogram to account for different staining processes. Sharma et al. [7] used conventional classifiers and pre-trained VGG16, VGG19 and ResNet50 as the classifiers. The best model was VGG16 model with linear Support Vector Machine (SVM), which provided the accuracies of 93.97%, 92.92%, 91.23% and 91.79% at magnification levels of 40×, 100×, 200× and 400×, respectively. Agaba et al. [8] used Deep Neural Network (DNN) classifiers with four dense layers and claimed that their proposed method achieved exceptional high accuracies of 97.89%, 97.6%, 96.10% and 96.84% at magnification levels of 40×, 100×, 200× and 400×, respectively. Patil et al. [9] proposed a multi-classification framework based on the pre-trained VGG16, VGG19, ResNet34, ResNet50, ResNet101, Resnet152, DenceNet121 and DenseNet201. With fine-tuning adaptations on the magnification-dependent classifiers, they claimed that the DenseNet121 model achieved the highest accuracies of 89%, and 84% at magnification levels of 40× and 400×, respectively, and that the DenseNet201 model had the best accuracies of 88% and 81% at magnification levels of 100× and 200×, respectively. When all images at different magnification levels were combined, the best magnification-independent classifier was the DenseNet121 model, which achieved an accuracy of 86.54%.

Although the deep learning models are significant in achieving high accuracy, image preprocessing is also equally important. Histopathological images often exhibit variations due to different digitization scanners and inconsistent Hematoxylin and Eosin (H&E) staining processes. These variations in color, brightness, and contrast can degrade the performance of mitosis detection models [10]. Conventional normalization techniques run the risk of amplifying existing noise in images if applied directly. Gupta et al. [11] investigated and found that using grayscale transformation as a stain normalization approach reduced the mitosis detection accuracy. Kausar et al. [10] adopted an alternative approach, where images were initially converted into the LAB color space. Then, each of the LAB channels underwent normalization based on the mean and standard deviation of both source and reference images. Finally, the images were converted back to the RGB color space, resulting in a more effective normalization method. However, this method was not used for breast cancer classification.

In summary, significant progress has been made in prior breast cancer (BC) classification research utilizing the BreakHis dataset, leading to advancements in automated detection methods and improving diagnostic accuracy. However, current research gaps include: (1) From a clinical perspective, breast tumor type classification holds a higher urgency than simply detecting the presence of breast cancer, yet most prior research has predominantly focused on binary classification rather than exploring the complexities of multi-classification. (2) Most pathological images in clinical practice vary in magnification factors, and previous multi-classification methods mostly leaned towards magnification-dependent classification rather than magnification-independent classification. However, a feature extraction and tumor classification model suitable for a specific magnification factor may not be applicable to another magnification factor. (3) There is a lack of systematic studies on how to improve the BC classification accuracy through image preprocessing of histopathological images.

To address these challenges, this study contributes significantly to breast cancer (BC) classification using deep learning. First, a magnification-independent deep learning framework is established, enabling both binary and multi-class BC classifications with improved accuracy, offering a scalable solution across various magnification levels. Second, the study introduces the z-score normalization procedure for histopathological images, which has not been mentioned in previous literature on BC classification. This technique enhances consistency and reliability across different datasets. Finally, the impact of dataset balance on classification performance is examined by comparing models trained on balanced versus unbalanced datasets, with data augmentation applied to assess its influence on accuracy and generalizability. These findings not only address current gaps in the literature but also provide actionable insights that can inform the development of more effective diagnostic tools for BC.

III. Methodology

A. BreakHist Dataset

As shown in Table I, BreakHis dataset [12] used in this study consists of 7,909 microscopic images of breast tumor tissue collected from 82 patients at various magnification factors, 40×, 100×, 200× and 400×. The dataset is divided into 8 tumor types, as shown in Fig. 1. Four types of benign breast tumors are adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA); and four types of malignant tumors (breast cancer) are ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC).

TABLE I. BREAKHIS DATASET

Binary Type	Tumor Types	Magnification				Total	# of Patients
		40x	100x	200x	400x		
Benign	Adenosis (A)	114	113	111	106	444	24
	Fibroadenoma (F)	253	260	264	237	1014	
	Phyllodes tumor (PT)	109	121	108	115	453	
	Tubular adenoma (TA)	149	150	140	130	569	
Malignant	Ductal Carcinoma (DC)	864	903	896	788	3451	58
	Lobular carcinoma (LC)	156	170	163	137	626	
	Mucinous carcinoma (MC)	205	222	196	169	792	
	Papillary carcinoma (PC)	145	142	135	138	560	
	Total	1995	2081	2013	1820	7909	

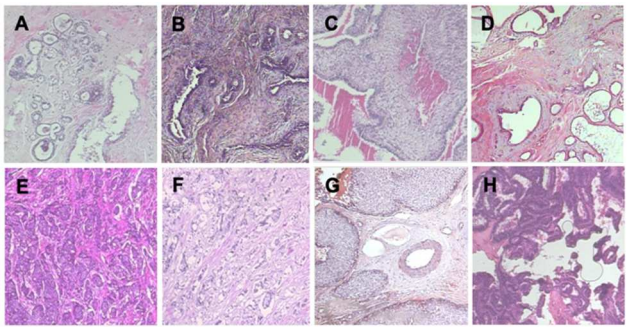

Fig. 1. The eight breast tumor types in the BreakHis dataset: Benign, (a) adenosis, (b) fibroadenoma, (c) phyllodes tumor, (d) tubular adenoma; Malignant, (e) ductal carcinoma, (f) lobular carcinoma, (g) mucinous carcinoma, (h) papillary carcinoma.

B. *Image Preprocessing and Classification Flow*

The dataset was utilized to create an automatic breast cancer classification model, and its workflow is presented in Fig. 2. First, all histopathology images in BreakHis dataset, initially sized at 700×460×3, were resized to a resolution of 224×224×3. The dataset was then partitioned into 70% for training and 30% for testing. Since the data is imbalanced, the deep learning models' classification performance before and after data augmentation was investigated and compared. To address the imbalance, SMOTE (Synthetic Minority Over-sampling Technique) library was utilized. This involves oversampling the minority class images by randomly applying horizontal flipping, vertical flipping, and rotation, ensuring that all seven minority classes are augmented to match the size of the majority subclass. This augmentation process increases the training set from 5536 to 19323 samples, with each subclass containing 2404 samples. It's worth noting that regardless of whether data augmentation is applied, the normalization process outlined in the following subsection is implemented before training begins.

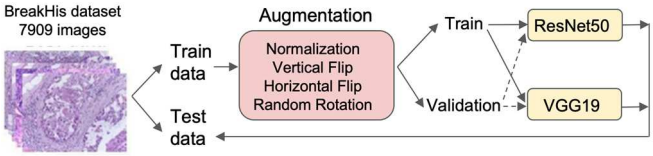

Fig. 2 Breast Cancer Detection and Tumor Type Classification Flow

C. *RGB Z-Score Normalization (or Standardization)*

In contrast to feature extraction techniques mentioned in [6][8], this study utilized color z-score normalization, which is different from the normalization methods in [10][11]. When the images are converted into PyTorch tensors, the original pixel range of [0, 255] is scaled to a PyTorch tensor range of [0.0, 1.0]. The means and standard deviations of the RGB channels across all training images are computed and utilized for z-score normalization, as shown in Equation (1).

$$x' = \frac{x-\mu}{\sigma} \quad (1)$$

where μ is the mean and σ is the standard deviation of all training images in RGB channels. x is the original pixel value and x' is the new pixel value after normalization. Following normalization, the new means and standard deviations of the RGB channels in all training images are set to 0.0 and 1.0, respectively. This normalization process effectively minimizes color and brightness variations, contributing to enhanced training robustness and improved accuracy.

D. *Convolutional Neural Network and Transfer Learning*

In this study, the pre-trained VGG16 was first utilized, but the performance was not satisfied. ResNet50 and VGG19 were later selected for the classification of breast cancers from the histopathological images due to their remarkable performance in image classification, object detection, and image segmentation.

The ResNet50 architecture [13], shown in Fig. 3(a), is composed of 50 layers, including multiple convolutional layers, activation layers, including residual blocks. The residual blocks enable the network to learn residual functions, enhancing its capacity to learn and generalize effectively. The VGG19 architecture [14] shown in Fig. 3(b) comprises 19 layers, including 16 convolutional layers and 3 fully connected layers. Its design, characterized by the consistent use of 3 × 3 convolutional filters with a stride of 1 and max-pooling layers, allows for effective feature extraction and pattern recognition.

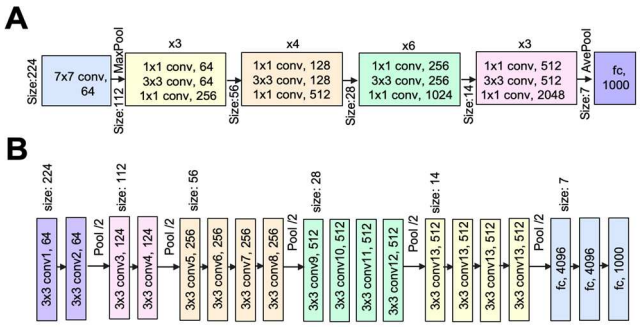

Fig. 3 Architectures of (a) ResNet50 and (b) VGG19

Among the training images, a random selection of 25% forms the validation set, which serves the dual purpose of performance evaluation and preserving optimal weight coefficients. The Adam optimization function is chosen for its superior accuracy and quicker convergence compared to SGD. Regularization is implemented with a weight decay of 0.0001, and categorical cross-entropy is employed for loss calculation. To ensure fairness in comparison, all models are trained with a fixed random seed. Throughout model training

process, a fixed batch size of 32 is maintained, with the epoch set to 30 in most cases, and occasionally 50. An adaptive learning rate strategy begins at 0.0003, gradually decreasing to 1e-5. Weights are saved, and the model is assessed on the validation set after each epoch. Upon completing the training, the model is tested on the test images to classify various breast tumor types. Data loading and normalization are facilitated using PyTorch 1.7.1, data splitting is managed with skorch-0.15.0, performance metrics are calculated using Scikit-learn, and data augmentation is handled with SMOTE. The training process is executed on Google Colab utilizing the V100 GPU hardware accelerator.

Both ReNet50 and VGG19 models are trained and evaluated with the dataset mixed with different magnification factors: 40×, 100×, 200× and 400×. To assess the overall performance of the proposed models, the trained models are tested using 2,373 images irrespective of magnification factors. Four metrics are used to evaluate the performance of the classification: accuracy, precision, recall (sensitivity) and F1 score. The confusion matrix is used to provide an intuitive understanding of the performance of a predictive model.

IV. Result and Discussion

Table II shows the performance metrics of breast cancer binary detection and 8-tumor multi-classification before and after data augmentation in this work. Precision, recall and F1-score are all the macro-average of classification result. Without data augmentation, ResNet50 and VGG19 models for binary classification achieved accuracies as high as 99.03% and 96.59%, respectively. Fig. 4(a) and 4(b) show the confusion matrices of the binary classification without data augmentation. These results indicate that the ResNet50 model exhibits higher reliability than VGG19 in capturing the distinctive features of benign and malignant tumors.

The performances of 8-tumor multi-classification with and without data augmentation were compared and discussed as follows. The accuracies of multi-classification for ResNet50 before and after data augmentation were 93.97% and 94.18%, respectively. This indicates that data augmentation did not significantly contribute to improving

TABLE II. PERFORMANCE METRICES OF BREAST CANCER DETECTION AND 8-TUMOR MULTI-CLASSIFICATION IN THIS WORK

Classification	Model	Accuracy	Precision	Recall	F1-score	Augmentatio
Binary	ResNet50	99.03%	98.53%	98.40%	98.46%	No
	VGG19	96.59%	96.00%	96.10%	96.05%	No
8-Class	ResNet50	93.97%	93.08%	93.45%	93.24%	No
	ResNet50	94.18%	93.11%	94.12%	93.53%	Yes
	VGG19	88.33%	86.95%	85.25%	85.97%	No
	VGG19	92.50%	91.46%	91.38%	91.26%	Yes

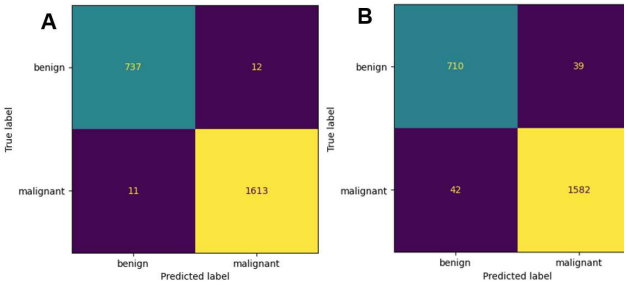

Fig. 4 Confusion Matrices of (a) ResNet50 and (b) VGG19 for binary classification without data augmentation.

the performance of ResNet50. On the other hand, data augmentation helped improve the accuracy of the VGG19 model from 88.33% to 92.50%. This improvement can be further explained by examining the learning curves. Fig. 5 shows the loss and accuracy learning curves of ResNet50 and VGG19 before data augmentation. In Fig. 5(a), when using the ResNet50 model without data augmentation, the training loss in blue and the validation curves in orange decreased together, with the final converged loss value showing only a 0.05 delta. This implies that the constructed ResNet50 framework was capable of distinguishing between the eight types of tumors accurately, even without data augmentation. However, Fig. 5(c) shows that at the end of the training process with the VGG19 model, there was a larger gap 0.2 between the training and validation loss without data augmentation. The final training and validation accuracies of

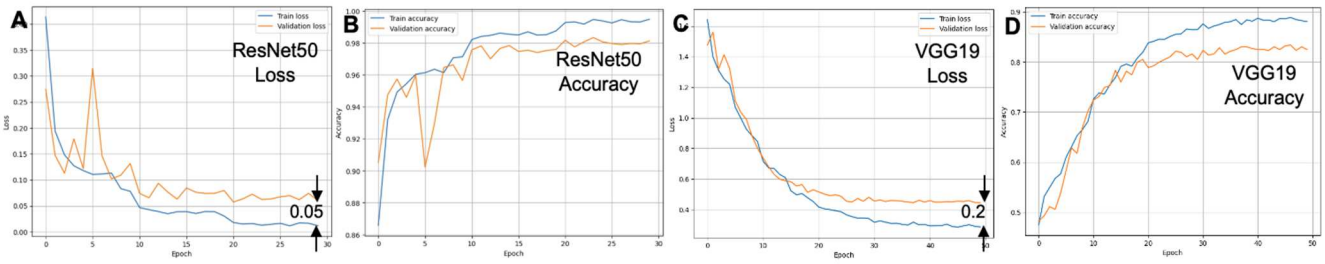

Fig. 5 Learning curves for multi-class classification without augmentation (a) ResNet50 loss (b) ResNet50 accuracy (c) VGG19 loss (d) VGG19 accuracy

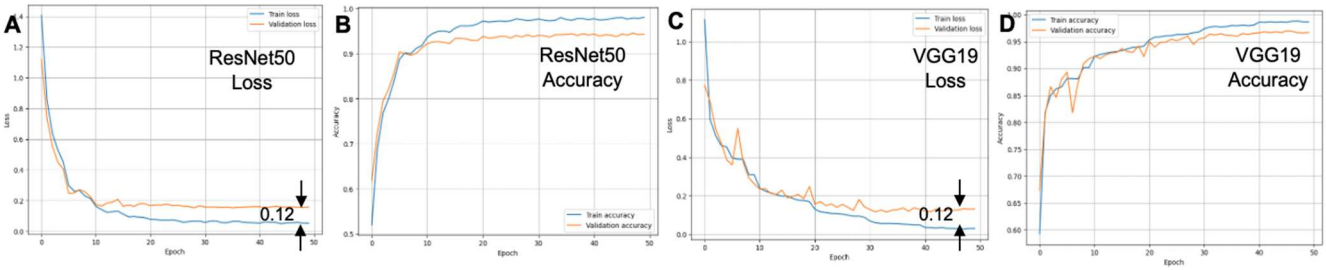

Fig. 6 Learning curves for multi-class classification with data augmentation (a) ResNet50 loss (b) ResNet50 accuracy (c) VGG19 loss (d) VGG19 accuracy

VGG19 in Fig. 5(d) are also lower than those of ResNet50 in Fig. 5(b).

Fig. 6 shows the learning curves for ResNet50 and VGG19 models after data augmentation. It is noticeable that data augmentation helped the learning curves converge more steadily. Fig. 6(c) and 6(d) illustrate that data augmentation is beneficial in addressing the overfitting concern in the VGG19 model, thereby narrowing the gap between validation loss and training loss and resulting in higher training, validation and test accuracies.

Fig. 7 and 8 show the confusion matrices of the ResNet50 and VGG19 models for BC multi-classification before and after data augmentation. Fig. 9 displays the performance metrics versus the eight BC tumor types. The findings in Fig. 7-9 reveal that the two most challenging tumor types to classify were ductal carcinoma (DC) and lobular carcinoma (LC). The similarity in patterns between DC and LC cells led to numerous misclassifications between these two types. Despite utilizing data augmentation to oversample all minority classes, the accuracy improvements for both ResNet50 and VGG19 models were limited. This may be attributed to the limited dataset, which was extracted from only 82 patients. The LC class has only 626 images, which was insufficient to build a robust classifier for this category. Notably, in the VGG19 results, many PT cells were misclassified as F, and many DC cells were misclassified as MC, whereas ResNet50 demonstrated a greater ability to capture the features of these cells, resulting in higher accuracy.

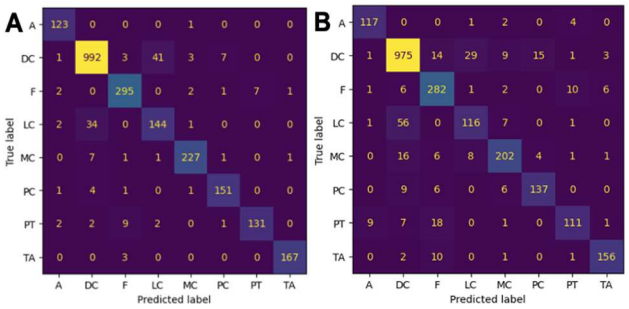

Fig. 7 Confusion matrices of multi-classification before data augmentation (a) ResNet50 Accuracy = 93.97%; (b) VGG19: Accuracy = 88.33%.

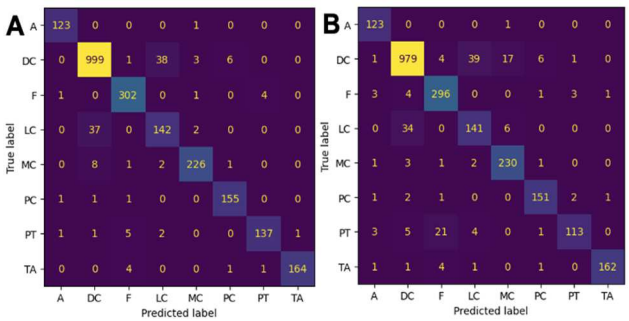

Fig. 8 Confusion matrices of multi-classification after data augmentation (a) ResNet50 Accuracy = 94.73%; (b)VGG19: Accuracy = 92.85%.

A. Comparison of Performance with Existing Research

Table III shows a comparison of the performance between our established ResNet50 and VGG19 deep learning breast cancer classifiers and the previous research in the most recent 5 years utilizing the BreakHis dataset. Many previous studies separated the images by the magnification factor first to reduce microscopy image variability and focused on developing magnification-dependent feature extractors and tumor classifiers. This paper focused on the developments of magnification-independent breast cancer classifiers.

TABLE III. CLASSIFICATION PERFORMANCE COMPARISON WITH PREVIOUS RESEARCH USING BREAKHIS DATASET

Ref	Year	Model	Magnification	Accuracy	Precision	Recall	F1-Score
Binary Classification							
[6]	2023	DenseNet169	Dependent (200x and 400x)	87.14%	91.21%	90.54%	78.93%
		Bit-S-R101x1		86.88%	90.67%	90.79%	77.45%
[5]	2023	VGG19	Independent	96.46%	93.14%	95.77%	94.43%
This Work	2024	ResNet50	Independent	99.03%	98.90%	98.86%	98.88%
		VGG19	Independent	96.59%	96.00%	96.10%	96.05%
Mult-class Classification							
[7]	2020	Handcrafted Features VGG16+ SVM	Dependent (40x)	93.97%	94.00%	93.00%	94.00%
			Dependent (200x)	91.23%	92.00%	92.00%	92.00%
[8]	2022	Handcrafted Features + DNN	Dependent (40x)	97.89%	97.00%	98.00%	97.00%
[9]	2022	DenseNet121	Dependent (40x)	89.00%	90.00%	85.00%	87.00%
		VGG19	Independent	68.00%	-	-	-
		ResNet50	Independent	80.00%	-	-	-
		DenseNet121	Independent	87.20%	85.00%	83.00%	87.00%
This Work	2024	ResNet50	Independent	94.73%	94.30%	94.23%	94.25%
		VGG19	Independent	92.50%	91.46%	91.38%	91.26%

When comparing the magnification-independent binary classifiers, the established VGG19 framework reached the similar accuracies but lightly higher precision, recall and F1-score by 3%, 0.3% and 2.5%, respectively, than the VGG19 classifier in [5]. The performance improvement could be due to the image preprocessing using Z-score normalization in RGB channels, which helped reduce the color and brightness variations. The proposed ResNet50 classifier of this study achieved 99.03% accuracy, outperforming state-of-the-art methods utilizing the same dataset.

When comparing the magnification-independent muti-classification, the established ResNet50 and VGG19 frameworks had higher accuracies, 94.73% and 92.5%, respectively, than the accuracies of 87%, 80% and 68% using DenseNet121, Resnet50 and VGG19 models, respectively, in [9]. Image normalization was mentioned in [9], but there was no explanation how the normalization process was computed. We believe our z-score normalization was the key factor for the performance improvement.

The established ResNet50 and VGG19 deep learning models were proven to be robust through the validation in the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 Blindness Detection dataset [15] which comprises 3,662 retinal fundus images and remarkable accuracies were achieved for diabetic retinopathy severity classification [16].

It should be noted that the proposed deep learning frameworks have several limitations: (1) The BreakHis dataset was extracted from only 82 patients with an imbalanced number of images. Several tumor categories have only a couple hundred images, which is insufficient to build a robust classifier; (2) The current framework is limited to the pre-trained ResNet50 and VGG19 models; and (3) The current model performance is based on the BreakHis dataset. Future work includes: (1) Increasing the training set size by merging with other dataset or finding a larger and balanced dataset; (2) Exploring other pre-trained models, such as

ResNet-RS, EfficientNetV2, and Vision Transformers; and (3) Examining more effective image preprocessing methods to further improve the performance.

V. Conclusion

Prior research using the histopathology images in the BreakHis dataset for breast cancer classifications mostly separated the images into different magnification factors before applying deep learning models for breast cancer detection and tumor type classification. In this paper, the established transfer learning-based deep learning models are not susceptible to variations in magnification factors. The Z-score image normalization played a key role in differentiating our approach from prior research, reducing color and brightness variation, and improving deep learning model performance. The results demonstrate that ResNet50 is a well-fitted model, with training loss and validation loss in proximity throughout the training process, even without data augmentation. ResNet50 achieved exceptional high accuracies of 99.03%, 93.97% and 94.73% for binary and multi-classification before and after augmentation, respectively. Data augmentation is beneficial in addressing the slight overfitting issue in VGG19, narrowing the gap between validation loss and training loss. VGG19 achieved accuracies of 96.59%, 88.33% and 92.50% for binary and multi-classification before and after augmentation, respectively. The proposed frameworks demonstrate high success in early breast cancer detection and eight-tumor-type classification, with better accuracy than other existing magnification-independent breast cancer classifier.

References

[1] International Agency for Research on Cancer, World Health Organization, available online: (accessed on 1 Aug 2023) https://gco.iarc.fr/today/en/dataviz/pie?mode=cancer&group_populations=1&types=0&sexes=2

[2] Ukwuoma, C.C.; Hossain, M.A.; Jackson, J.K.; Nneji, G.U.; Monday, H.N.; Qin, Z. Multi-Classification of Breast Cancer Lesions in Histopathological Images Using DEEP_Pachi: Multiple Self-Attention Head. *Diagnostics* 2022, *12*, 1152.

[3] A. Ibrahim, A. Lashen, M. Toss, R Mihai, E. Rakha, "Assessment of mitotic activity in breast cancer: revisited in the digital pathology era," *Journal of Clinical Pathology* 2022;75**:**365-372.

[4] A. Kanavos and P. Mylonas, "Deep learning analysis of histopathology images for breast cancer detection: a comparative study of ResNet and VGG architectures," *2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP 2023)*, Limassol, Cyprus, 2023, pp. 1-6.

[5] R. Mahadeva, S. P. Patole, V. Patel, V. Chaurasia, A. Sharma and R. Sharma, "Deep transfer learning with multi-level features extraction approach for breast cancer classification," *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*, BHOPAL, India, 2023, pp. 471-474

[6] D. F. Fauzan, R. Fauzi, O. N. Pratiwi and J. M. Ferreira Machado, "Breast cancer detection on histopathology images using pre-trained computer vision models," *2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Bali, Indonesia, 2023, pp. 1-6.

[7] S. Sharma and R. Mehra, "Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image," The Visual Computer, 36(9), 2020, pp.1755-1769

[8] A.J. Agaba, M. Abdullahi, S.B. Junaidu, H.I. Hayatu, and H. Chiroma, "Improved Multi-Classification of Breast Cancer Histopathological Images using Handcrafted Features and Deep Neural Network (Dense layer)", Intelligent Systems with Applications, 2022, p. 200066.

[9] N. S. Patil, S. D. Desai and S. Kulkarni, "Magnification independent fine-tuned transfer learning adaptation for multi-classification of breast cancer in histopathology images," *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 2022, pp. 1185-1191

[10] T. Kausar, M. Wang, B. Wu, M. Idrees and B. Kanwal, "Multi-scale deep neural network for mitosis detection in histological images," *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Bangkok, Thailand, 2018, pp. 47-51

[11] V. Gupta, A. Singh, K. Sharma, and A. Bhavsar, "Automated classification for breast cancer histopathology images: Is stain normalization important?" in Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures—CARE (Lecture Notes in Computer Science), vol. 10550, M. Cardoso et al., Eds. Cham, Switzerland: Springer, 2017, available online (accessed on 12 Dec 2023): https://link.springer.com/chapter/10.1007/978- 3-319- 67543-5 16.

[12] Kaggle Breast Cancer Histopathological Database (BreakHis), Available online: https://www.kaggle.com/datasets/ambarish/breakhis (accessed on 12 July 2023.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 770–778.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society, 2015, pp. 1–14.

[15] APTOS 2019 Blindness Detection, Available online: https://www.kaggle.com/c/aptos2019-blindness-detection (accessed on 13 November 2022).

[16] N. Nishio, J. Nishio, and A. Nishio, "Early Detection and Severity Classification of Diabetic Retinopathy Through Image Processing and Deep Learning," 2024 IEEE International Conference on Future Machine Learning and Data Science, Sydney. in press.