IMPROVED BREAST CANCER DIAGNOSIS THROUGH TRANSFER LEARNING ON HEMATOXYLIN AND EOSIN STAINED HISTOLOGY IMAGES

A PREPRINT

⊙ Fahad Ahmed
TIB - Leibniz Information Center for Science and Technology, Hannover, Germany
CARESAI
fahad.ahmed@tib.eu

Reem Abdel-Salam
Cairo University, Faculty of Engineering, Computer Department, Egypt
CARESAI
reem.abdelsalam13@gmail.com

Leon Hamnett
CARESAI
L.s.hamnett@gmail.com

Mary Adewunmi
College of Health & Medicine(CHM), School of Medicine, University of Tasmania(UTAS), Hobart, Australia
CARESAI
mary.adewunmi@utas.edu.au

Temitope Ayano
Department of Microbiology, Obafemi Awolowo University, Ile-Ife, Nigeria
CARESAI
ayanotemitope@gmail.com

November 27, 2023

ABSTRACT

Breast cancer is one of the leading causes of death for women worldwide. Early screening is essential for early identification, but the chance of survival declines as the cancer progresses into advanced stages. For this study, the most recent BRACS dataset of histological (H&E) stained images was used to classify breast cancer tumours, which contains both the whole-slide images (WSI) and region-of-interest (ROI) images, however, for our study we have considered ROI images. We have experimented using different pre-trained deep learning models, such as Xception, EfficientNet, ResNet50, and InceptionResNet, pre-trained on the ImageNet weights. We pre-processed the BRACS ROI along with image augmentation, upsampling, and dataset split strategies. For the default dataset split, the best results were obtained by ResNet50 achieving 66% f1-score. For the custom dataset split, the best results were obtained by performing upsampling and image augmentation which results in 96.2% f1-score. Our second approach also reduced the number of false positive and false negative classifications to less than 3% for each class. We believe that our study significantly impacts the early diagnosis and identification of breast cancer tumors and their subtypes, especially atypical and malignant tumors, thus improving patient outcomes and reducing patient mortality rates. Overall, this study has primarily focused on identifying seven (7) breast cancer tumor subtypes, and we believe that the experimental models can be fine-tuned further to generalize over previous breast cancer histology datasets as well.

1 Introduction

Breast cancer is one of the major causes of death among women worldwide [37, 11]. Early identification can minimize fatality rates, thereby, preventing disease mutation and progression [38]. A pathologist examine tissue biopsy to confirm the presence of cancerous cells in breast [3]. The tissue is fixed to slides and examined in a lab after the biopsy is finished. Formalin fixation is the first step in the tissue preparation procedure, followed by embedding in paraffin sections [2]. The paraffin blocks are subsequently cut into pieces and mounted on glass slides. However, the tissue's cytoplasm and nuclei, which contain valuable information for the pathologist, are not yet visible.

Hematoxylin and eosin (H&E) staining [10], a common and well-known staining procedure for tissue visibility. Hematoxylin can bind to deoxyribonucleic acid when it is added to the tissue, which causes the cell nuclei to become stained blue or purple. The stroma and cytoplasm, on the other hand, are colored pink because eosin has the ability to bind to proteins and dye these structures. The glass slide is typically coverslipped and set in place after staining and sent to a pathologist for review. The expert routinely collects data on the nuclei's size, shape, organization, interactions, and spatial configurations. The variability, density, and general structure of the tissue are also examined. For identifying carcinoma (a type of cancer that forms in epithelial) and non-carcinoma cells, information on the characteristics of the nuclei, such as shape and density is particularly important. Additionally, knowledge of tissue structure is important for identifying benign, atypical, and malignant classes of cells.

In this study, we used the up-to-date BRACS dataset [1] to develop our image classification models of histopathological (H&E) stained images. The ductal carcinoma in situ (DCIS) and invasive carcinoma (IC) lesions are malignant tissues that need immediate medical attention because if left untreated, they will continue to progress and will cause a fatality in the patient. In-situ carcinoma refers specifically to the development of atypical cells that are restricted to the layer of breast tissue from which they originated. Atypical cells that have invaded normal tissue outside of the glands or ducts from which they originated are referred to as invasive carcinoma.

Treatment for invasive carcinoma is challenging because the cancer cells may spread to other organs [39] and the treatment can put the entire body at risk. As the cancer progresses through more advanced stages, the likelihood of survival decreases. Additionally, a patient's in-situ carcinoma tissue may transform into invasive carcinoma tissue in the absence of appropriate and effective treatment [7]. Therefore, it is crucial to effectively examine the biopsy tissue so that a diagnosis can be made and treatment can start. It is not an easy or simple task to conduct a practical examination of the tissue. Instead, it takes a lot of time and is, more importantly, prone to human error. Professionals have a diagnostic accuracy rate of about 75% on average [47]. Patients with incorrect diagnoses could suffer serious and even fatal consequences as a result of these problems.

The need for a system that can automatically and accurately classify breast cancer histopathology images is generally justified by the lack of trained medical specialists [48] based on the population of patients they must serve, as well as the laborious process of arriving at a final diagnosis, and the problem of interobserver variability.

While providing adequate classification accuracies based on data availability, previous research solutions have been computationally expensive. Convolutional neural networks (CNNs) have used multiple instance learning (MIL) to classify cancer histopathology images. The WSI is divided into patches for gigapixel classification. MIL patches are instances in bags. Multiple instances learning is called that because each image (WSI) has many patches, the bag contains several instances representing different patches. Training supervision requires only global picture labels. To make a prediction, the aggregation process summarises all instance data. These multiple instance learning (MIL) models are based on vision transformer models, which are computationally expensive and require large image datasets for effective training.

Patch-based image classification is difficult to apply to high-resolution breast cancer histopathology. Due to large image sizes and high detail, WSIs require a different approach to gain insights. Patch selection approaches are needed because WSIs are divided into patches and cannot fit into memory when training a machine-learning model. If class labels are only available on a WSI basis, it may be difficult to determine which patches contain critical information. No global representative field exists, making model decisions hard to understand and trust.

Due to the above reasons, image classification models that use less memory and require smaller image datasets are needed. Lightweight classification models that perform like current deep learning approaches are essential. This study poses the following main research question: "Can tuned Exception, EfficientNet, ResNet50, ConvNextTiny, and Inception-ResNet achieve similar results to current state-of-the-art MIL approaches for classifying gigapixel images of breast cancer histology WSIs?" . Additionally, we explore related questions such as: What pre-trained image classification models are suitable for this task? Can normal patch-based transfer learning work? What data augmentation approaches could improve breast cancer histology slide classification?

This study classifies breast cancer histology gigapixel images using pre-trained image classification models like Xception, EfficientNet, Resnet50, and InceptionResNet, as well as pre-processing techniques like resizes, tiling, stain normalisation, and image compression to improve classification performance. We also use transfer learning to improve classification performance by adapting models from large image datasets to our task. To make models more generalizable and robust to image dataset characteristics, we emphasize augmentation techniques like flipping rotation.

The rest of the paper is organized as follows: In section 2, Literature Review describes previous related research work and successful strategies. Section 3, Methodology and Techniques, describes this study's framework. Section 4, Discussion and Results details the research and experimentation findings. Section 5 summarizes the paper's findings.

2   Literature Review

Breast cancer is a significant public health concern globally, with an estimated 2.3 million new cases identified in 2020 alone [27]. Early detection of breast cancer is vital for improved patient outcomes, as treatment success often depends on the stage at which the cancer is diagnosed. Although mammography is the most widely used screening method for breast cancer detection, it has some limitations, particularly in women with dense breast tissue [28]. These limitations include a relatively high false-positive rate, limited sensitivity in dense breast tissue, and the potential risk of radiation exposure due to repeated screening examinations [28].

Histopathology examination of tissue samples remains the gold standard for breast cancer diagnosis, as it analyzes the image at the cellular level, which is necessary to determine the malignancy of the tumor [29]. However, this method requires time-consuming, labor-intensive procedures subject to inter-observer variability. In recent years, deep learning algorithms have shown promising results in automating medical imaging analysis, including histology images, for breast cancer diagnosis.The literature review aims to provide an overview of the current state-of-the-art in deep learning-based approaches for breast cancer diagnosis using histology images.

There has been a growing interest in developing computer-aided diagnosis (CAD) systems for breast cancer using histology images. To address the limitations of previous approaches for multi-classification of histology images, [32] explored the use of handcrafted feature extraction techniques, including Hu moments, Haralick textures, color histograms, and a deep neural network (DNN), for breast cancer multi-classification on the BreakHis dataset. This method performed better in the multi-class classification of breast cancer than most of the related works in the literature. However, [26] compared two machine-learning methods for automatically classifying breast cancer histology images into benign and malignant subclasses. The authors tested dataset augmentation techniques to improve the convolutional neural network's accuracy and found that it outperformed the handcrafted feature-based classifier, achieving accuracy between 96.15% and 98.33% for binary classification and 83.31% and 88.23% for multi-class classification. To address class imbalance in classifying mitotic and non-mitotic nuclei in breast cancer histopathology images.

[25] proposed a CNN architecture and data balancing technique that yielded an F-measure of 0.79, outperforming all methods relying on handcrafted features and those using a combination of handcrafted and CNN-generated features. The results from these studies suggest that CNN-based approaches outperform those relying on handcrafted features, with dataset augmentation techniques and data balancing techniques showing potential for further improving classification accuracy.

Several studies have proposed CNN-based approaches for breast cancer histology image classification. [31] proposed a method that can extract information from relevant scales, including nuclei and overall tissue organization, achieving an overall sensitivity of approximately 81% for carcinoma patch-wise classification. This proposed system can also be extended to whole-slide breast histology image classification relevant to clinical settings.

[24] also proposed a CNN-based CAD system that assesses tumor-associated stroma in hematoxylin and eosin-stained breast specimens. Their method achieves high accuracy for the binary classification of breast tissue biopsies. A class structure-based deep convolutional neural network (CSDCNN) model with multiple hidden layers that achieved high accuracy by preserving intra-class variance using a novel distance constraint for feature space was developed by [30], and a capsule network with dynamic routing to classify breast cancer tissue images that achieved promising results for different classes of breast cancer tissue images has also been proposed by [23].

Furthermore, several studies have proposed novel deep CNN architectures for breast cancer histology image classification. [22] proposed a novel deep CNN and advanced data augmentation methods, achieving higher classification accuracy (up to 97%) with good robustness and generalization. [21] proposed a novel CNN with a small SE-ResNet module and a new learning rate scheduler for automatic classification of histology images into benign and malignant and eight subtypes with high accuracy.

[20] proposed a fine-tuning strategy for ResNet for the multi-classification of breast cancer histology images into normal tissue, benign lesions, in-situ carcinomas, and invasive carcinomas, performing remarkably well on images provided for the Grand Challenge on Breast Cancer Histology Images (BACH) [40]. [15] proposed a hybrid convolutional and recurrent deep neural network for breast cancer histopathological image classification. Their model achieved an impressive 95.75% classification accuracy.

[19] utilized a patch-based deep learning technique consisting of two consecutive convolutional neural networks to achieve 95% accuracy on the validation set. [33] proposed a computational approach based on deep convolutional neural networks for breast cancer histology image classification, achieving an accuracy of 87.2% and 93.8% for the 4-class and 2-class classification tasks, respectively.

[18] proposed a residual architecture model with attention modules and augmentation techniques called BreastNet, achieving stable and accurate classification using the hypercolumn technique, achieving a 98.80% classification. [13] modified the convolutional and fully connected layers of a pre-trained AlexNet, achieving patch and image-wise accuracy of 75.73% and 81.25%, respectively, and image-wise accuracy of 57% on the ICIAR-2018 breast cancer challenge hidden test set.

In a study by [14], transfer learning and fully trained networks for the multi-classification of breast cancer were compared. They found that a fine-tuned, pre-trained VGG16 with a logistic regression classifier had the best performance, achieving 92.60% accuracy, a 95.65% area under the ROC curve (AUC), and a 95.95% accuracy precision score (APS) [14]. [41] utilized Inception V3 for patch-level classification and a Dual Path Network (DPN) as a feature extractor for multi-classification of histological images. Their approach achieved a 12.5% improvement over the state-of-the-art model in experimental results.

Finally, [17] designed a 152-layered CNN named ResHist to classify histopathological images into benign and malignant classes using their data augmentation technique. High-throughput Adaptive Sampling for Whole-Slide Histopathology Image Analysis (HASHI) was proposed by [16] to detect invasive breast cancer on whole-slide histopathology images (WSI). HASHI uses a CNN-based robust representation learning classifier and a new adaptive sampling method based on probability gradients and quasi-Monte Carlo sampling. Transfer learning has improved breast cancer diagnosis speed and efficiency. Researchers have used pre-trained deep convolutional neural networks (CNNs) to improve breast cancer image classification performance as this enables models to transfer knowledge from domains with substantial image data to domains with limited data.

3 Methodology and Techniques

In this research study, we focus on two types of experiments. The first type of experiment involves applying transfer learning on the reduced-resolution images since the original images are extremely high-resolution and also investigating the impact of transfer learning on high-resolution images. In the second type of experiment, we explore the application of transfer learning using image tiling applied to lower-quality and higher-quality images. The dataset used for this project, its statistics, the pre-processing steps taken, the data augmentation techniques, and the training models are all discussed in this section.

3.1 Dataset

BRACS [1] is a set of histopathological slides with hematoxylin and eosin(H&E) staining applied. It contains atypical lesions as well as six different subtypes of lesions. This also contains histological normal tissue samples. The BRACS dataset is created to support the development of breast cancer diagnostic methods through the automated analysis of histology images. The dataset was developed through the collaboration of the National Cancer Institute - Scientific Institute for Research, (IRCCS) "Fondazione G. Pascale", the Institute for High-Performance Computing and Networking (ICAR) of National Research Council (CNR), and IBM Research – Zurich. The dataset was acquired from patients between 2019 and 2020, by board-certified pathologists of the Department of Pathology at the National Cancer Institute- IRCCS "Fondazione G. Pascale" in Naples (Italy). The dataset used in this study was made available to the public in 2021 [1].

3.2 Dataset Statistics

The samples were generated from H&E-stained breast tissue biopsy slides and were selected based on the diagnostic reports of the patients. The age of the patients ranges from 16 to 86 years, with about 61% of patients in the range of 40-60 years, and only a small number of patients being aged less than 20 years or above 80 years. Due to the laborious and expensive process of data collection, datasets in the medical domain typically consist of relatively small numbers

of images. The dataset includes 4539 Regions of Interest (ROIs) that were extracted from 547 Whole-Slide Images (WSIs). Three lesion types are present within the BRACS dataset: benign, malignant, and atypical. These three lesion types are further subtyped into seven groups: Normal (N), Pathological Benign (PB), Usual Ductal Hyperplasia (UDH), Flat Epithelial Atypia (FEA), Atypical Ductal Hyperplasia (ADH), Ductal Carcinoma in Situ (DCIS), and Invasive Carcinoma (IC). The included ROIs have a wide range of image sizes and incorporate common visual artifacts from the tissue-preparation and staining process, that are typical of the types of patterns that would be seen in a clinical setting or pathological laboratory that makes diagnosis more challenging. A total of 7909 breast cancer histopathology image samples, obtained from 82 patients at four different magnification levels, are included in the dataset. The entire dataset takes up about 52 GB of space. The BRACS dataset is a gigapixel dataset (high resolution) comprised of 4391 ROIs obtained from 325 H&E stained breast cancer WSI with varied sizes and appearances (at magnification level 0.25 µm/pixel for 40x resolution), which can easily exceed 100,000 by 100,000 pixels. The samples in the dataset are separated into two main groups: benign and malignant, with the benign group including 2480 samples and the malignant group including 5429 samples, respectively.

3.3 Preprocessing

For the classification of histology images, preprocessing is essential for the stability of model training and for better convergence[4]. In order to use BRACS dataset with conventional model training approaches, the resolution of the images must be reduced whilst still aiming to preserve important image features. To work around this issue we explore two paths. The first path: is to decrease the resolution of images by resizing and graying out the noise [9], which reduced storage space approximately to 404 MB. The other solution is to apply tiling and patching to the high-resolution images[8]. This involves splitting images into multiple smaller images, this retains the level of detail of the original images. The non-uniformity of BRACS ROI images makes it necessary to resize them to a standard image size as suggested by [5] as typical classification models are not able to accept a mix of image sizes. In our experiments, we focused on 5 different image dimensions: (300 x 300), (512 x 512), (256 x 256), (512 x 256), and (1024 x 1024).

3.4 Data augmentation

Data augmentation is a technique used to increase the amount of image variation in the training set because the dataset is much smaller than what is typically needed to train a deep-learning model correctly. This method helps prevent overfitting, a condition in which a model is overly influenced by the information in the training images and is not able to generalize and correctly classify new unseen images. Eight different types of image augmentations are used as data augmentation techniques for both the experiments relating to training models from initial conditions and the experiments involving transfer learning. These image augmentations can be classified as flip (reversing left-right or up-down orientation), rotating, brightening and darkening, resizing, cropping, blurring and sharpening, distorting, noise, and special (emboss, perspective, cutmix, mixup, etc.). A variety of techniques from the Albumentations and MONAI libraries were analysed to see how they contributed to increasing classification accuracy and improving model performance. The process of analyzing histology images is rotationally invariant, as a pathologist can view a microscopy image from any angle and still be able to successfully examine it. As a result, adding a rotation augmentation to the image shouldn't have a negative impact on classification performance. The rotation augmentation is tailored so that an image is rotated 90*k° in a clockwise direction, where k=0,1,2,3. The processed images as discussed in the previous section also received a random width and height shift (translating an image in a certain direction), zoom range (zooming in or out of an image), and horizontal and vertical flips.

3.5 Data upsampling

In essence, BRACS is an imbalanced dataset, as shown in Table 1. The potential drawbacks of class imbalance within a training dataset are well documented [42] and include a possible impact on the effectiveness of the model to distinguish between different classes. If the model is not shown a sufficient number of images corresponding to a certain class, it may not effectively learn image features relating to that class and may fail to correctly classify that type of image. Typically, models developed from unbalanced datasets perform poorly in terms of predicting outcomes for the minority class, leading to higher false-negative rates and lower recall [42].

Table 1: Original distribution of classes in BRACS dataset.

Class	N	DCIS	FEA	PB	IC	UDH	ADH
Samples	484	790	756	836	649	517	507

In order to mitigate the issue of unbalanced classes within the dataset, we applied different upsampling strategies such as batch-balanced sampler, weighted batch sampler, and general upsampling. In a batch-balanced sampler, we make sure that each batch contains an equal number of classes. In the weighted batch sampler, we give weight to each class according to its inverse probability of occurrence, then we sample images in each batch based on their class weights. Finally, in the general upsampling approach, the dataset is upsampled to approximately 1000 and 2000 samples for each class in order to obtain boost performance when classifying under-represented classes.

3.6 Modeling strategies

In this research study, we focus on two types of experiments. The first type of experiment involves applying transfer learning on the original images of high resolution as well as on the low-resolution versions. In the second type of experiment, we explore the application of transfer learning using image tiling applied to both high-resolution and lower-quality images.

3.6.1 Transfer learning

We implemented 5 well known deep learning classification models, all of which were pre-trained on the ImageNet dataset. These models include Xception, EfficientNet, ResNet50, ConvNextTinyV2, and InceptionResNet. These pre-trained CNN models are often employed in research tasks due to their significant performance against reference classification datasets such as ImageNet and CIFAR-100. Additionally, these models have shown relatively high performance for other challenging computer vision problems within varied research fields [43].

First, we apply the image pre-processing methodology to the dataset as discussed in the above section. We have investigated several loss functions such as cross-entropy loss function, focal loss, and cross-entropy with label smoothing with a smoothing value of 0.35. AdamW optimizer was used with a learning rate of 1e-3 and weight decay ranges from 1e-2 to 1e-5. In addition, a cosine annealing learning rate scheduler was used. In addition, we investigated the effects of different sampling strategies such as no sampling, weighted balanced sampler, and batch balanced sampler. In our experiments, we have investigated two model architectures.

The first architecture is to feed the image directly to the backbone model then the output is fed to the attention classifier layers as shown in figure 1. In addition, we inspect model performance using two methods original data split and custom data split.

The second architecture takes inspiration from pyramid networks [44]. At first, the images are fed to the backbone model. Then we extract the output activation of the last 4 layers. These outputs are then fed to convolution attention layers. Then those outputs are then fused and fed to a classifier head, as shown in figure 2. Both of these architectures are applied to high-resolution images and low-resolution images.
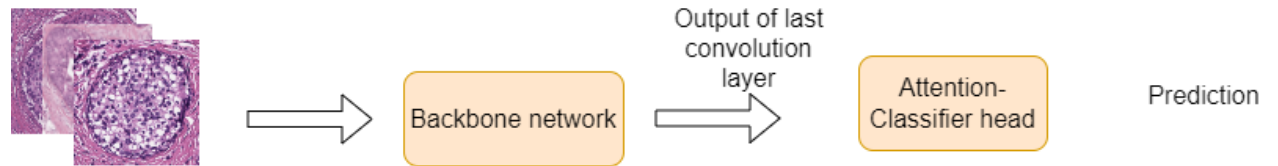

Figure 1: Breast histology image classification workflow by fine-tuning Xception, EfficientNet, ResNet50, ConvNext-TinyV2, and InceptionResNet network architectures.

3.6.2 Image tiling and transfer learning

In this approach, we combine both ideas of multi-instance learning and the standard fine-tuning process. First, for each image, we extract N tiles with different window sizes and zoom levels. Then we concatenate (merge) these tiles along the height or width dimension. This results in a large image size as it contains different tiles and variability. The resultant merged image is of size (1748,1748). The pipeline is shown in figure 3. In this experiment we have set the image size to (1024, 1024) and the number of tiles to 50, with windows of size 128,256,512 and 3 zoom levels.

4 Discussion and Results

We have used weighted f1-score, sensitivity, and accuracy as our main metrics to report performance on the BRACS test-set. In table 6, we compare the classification performance of different experiments when making predictions on the
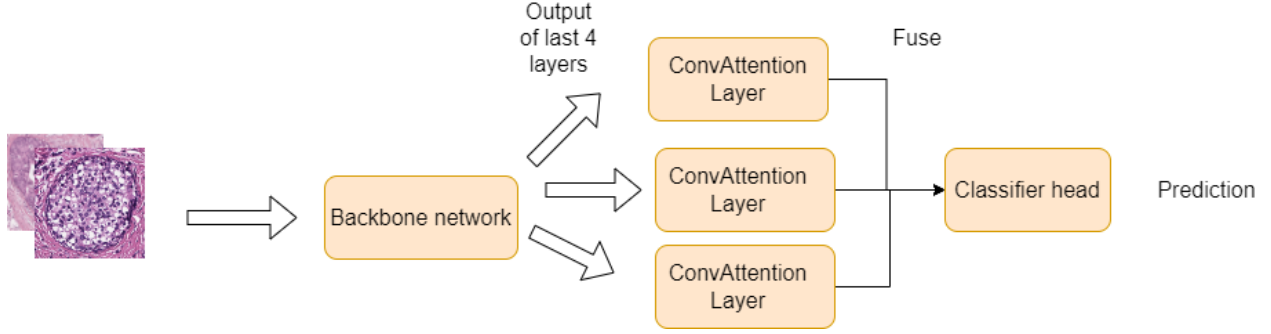
Figure 2: Breast histology image classification workflow by fine-tuning Xception, EfficientNet, ResNet50, ConvNext-TinyV2, and InceptionResNet pyramid network architectures.
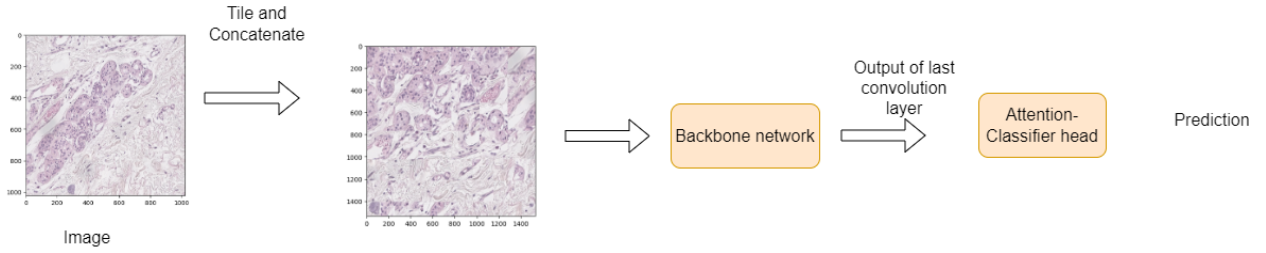

Figure 3: Breast histology image classification workflow by tiling and fine-tuning ResNet50 network.

test set provided in the BRCAS dataset, using the low-resolution images. The ResNet50 model with focal loss function and image size of (512,512) was able to achieve a high f1-score of 65. The ResNet50 pyramid architecture and the ResNet50 with cutmix and mixup augmentation both achieve the second-highest classification performance with an f1-score of 63. These metrics indicate that cutmix and mixup image augmentations, feature pyramids, and a focal loss function are key successes of high performance and training stability. In contrast to other models, ResNet50 showed promising results. Surprisingly, tiling-based methods didn't show high performance, this may need more investigation thoroughly to determine factors such as the best number of tiles to generate per image, as well as tile size, and whether tiles should contain overlapping areas.

When we investigated different sampling methods like weight-batch samplers and batch-balanced samplers and their impact on classification performance. They had almost yielded the same results with a 0.05 magnitude of difference. However, without them, performance degrades with an order of 10% or more in the f1-score. As seen in Figure 4, the ResNet50 model shows good performance when classifying N, FEA, and IC classes. However when making classification predictions, commonly, classes are mistaken for a different class entirely e,g ADH is mistaken for UDH or DCIS classes and DCIS is incorrectly predicted as ADH class.

In table 3, the performance of different classification models using a high-resolution dataset is reported on the test set provided in the BRCAS dataset. The ConveNextTiny v2 model was able to achieve high performance with a f1-score of 61. Pyramid networks did show promising results compared to tiling methods, order of 10% f1-score difference. This may lead to more investigation into pyramid models and different architectures.

In table 2, we compare results on the BRACS test dataset against state-of-the-art models such as ScoreNet and HACT-Net. Our ResNet50 model is fine-tuned on low resolution and shows good performance compared to SOTA models, with a magnitude of improvement of 10% in the f1-score. This highlights that another way to handle a gigapixel dataset problem is to use appropriate resizing and remove pixels thus creating lower resolution images, and to take advantage of pre-trained models which are using smaller image sizes to enable effective transfer learning.

Tables 4 and 5 show the results of different models using a custom dataset split. In this custom split, we have combined the train-validation-test images as given in the BRACS dataset and then randomly assigned each image into a data subset using a 0.9,0.07,0.03 probability split for training, validation, and test sets respectively. For images assigned to the training subset, we performed multiple experiments where each class was upsampled to 1000 or 2000 samples per class. During upsampling, EfficientNet was able to show good performance during both experiments. With 1000 samples per class, the model demonstrated a 0.77 f1 score. Additionally, with 2000 samples per class, the model showed

even greater performance with a 0.96 f1 score. This indicated that upsampling samples per image class can contribute significantly to better classification performance.
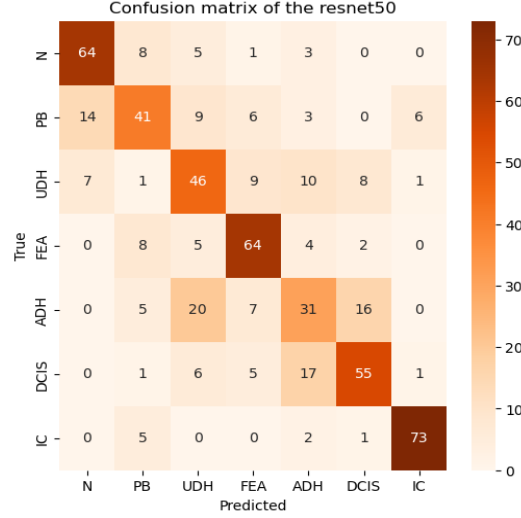

Figure 4: Confusion matrix of the predictions of the ResNet model (best-performing model) on BRACS test-set.

Table 2: Test results for the original BRACS dataset with the state-of-the-art models.

Model	F1-score
ResNet50 (Ours)	**0.65**
ScoreNet [36]	0.64
HACT-Net [35]	0.62

Table 3: Test results for the original BRACS dataset split using high-resolution images.

Model	Image Size	Special Augmentation/ Preprocessing Dropout rate	Loss Function	Accuracy	F1-score
ResNet50 Pyramid	(512,512)	CUTMIX and MIXUP	0.45	LabelSmoothing	0.59	0.59
ConvNextTiny V2	**(512,512)**	**NO**	**0.45**	**Focal Loss**	**0.61**	**0.6**
ResNet50	(512,512)	NO	0.45	Focal Loss	0.57	0.57
ResNet50 with Tiling	(1024,1024)	NO	0.45	Focal Loss	0.5	0.49

Table 4: Test results for the custom BRACS dataset split with 1000 upsampling.

Model	Accuracy	F1-score	Sensitivity
Xception	0.716	0.713	0.712
EfficientNet	**0.768**	**0.77**	**0.77**
ResNet50	0.73	0.725	0.729
InceptionResNet	0.725	0.725	0.724

8

Table 5: Test results for the custom BRACS dataset split with 2000 upsampling.

Model	Accuracy	F1-score	Sensitivity
Xception	0.884	0.884	0.887
EfficientNet	0.955	0.955	0.956
ResNet50	**0.962**	**0.962**	**0.962**
InceptionResNet	0.957	0.957	0.958

Table 6: Test results for the original BRACS dataset split using low-resolution images.

Model	Image Size	Special Augmentation/ Preprocessing	Dropout rate	Loss Function	Accuracy	F1-score
EfficientNet	(1024,512)	NO	0.35	Focal loss	0.59	0.59
EfficientNet	(512,512)	NO	0.45	Focal loss	0.62	0.62
EfficientNet	(512,512)	NO	0.5	Focal loss	0.51	0.49
ResNet50	(512,512)	NO	0.45	CrossEntropy	0.57	0.58
ResNet50	**(512,512)**	**NO**	**0.45**	**Focal Loss**	**0.66**	**0.65**
ResNet50	(512,512)	CUTMIX and MIXUP	0.45	LabelSmoothing	0.63	0.62
ResNet50	(512,512)	Stain Normalization	0.45	CrossEntropy	0.59	0.57
ResNet50 Pyramid	(512,512)	CUTMIX and MIXUP	0.45	LabelSmoothing	0.63	0.63
ResNet50 with Tiling	(1024,1024)	NO	0.45	Focal Loss	0.59	0.47
InceptionResNet	(512,512)	NO	0.45	Focal Loss	0.56	0.54

5 Challenges and Limitations

Different laboratories and institutions use different staining and imaging protocols for H&E stained histology images. Furthermore, tissue preparation, staining duration, reagents, and imaging equipment can affect image appearance and quality, making it hard to establish consistent patterns or features. Another issue is that pathologists or domain experts must manually label and classify histology images, which can lead to inter-observer variability and subjectivity, or even annotation inconsistencies between different reviewers. Also, different datasets may have quality assurance processes or labelling criteria, making it harder to establish consistent ground truth class labels across different image datasets. These obstacles prevent us from generalising our study across different datasets. However, we are excited to experiment on specific classes like Normal (N), Pathological Benign (PB), Usual Ductal Hyperplasia (UDH), Ductal Carcinoma in Situ (DCIS), and Invasive Carcinoma (IC) from the previous datasets.

6 Conclusion

During this research project, we aimed to answer one main research question: "Can tuned Xception, EfficientNet, ResNet50, ConvNextTiny, and Inception-ResNet models achieve comparable performance to current state-of-the-art multi-instance learning (MIL) approaches for the application of classifying breast cancer histology gigapixel images?" The research investigation that was undertaken showed that these models are suitable for the task of classifying images of breast cancer histology slides.

We performed three different approaches while undertaking this research. For the first approach, we experimented with the default BRACS dataset splits, and in our test results the ResNet50 model outperformed SOTA models achieving 65% F1-score. For our second approach, we split the dataset into custom train-val-test splits and also upsampled the dataset to 1000 samples per class, and in our test results the EfficientNet showed the best classification performance, achieving 77% F1-score and 77% sensitivity. For our third approach, we randomly split the dataset into train-val-test custom splits (90%-7%-3%) and also upsampled the dataset to 2000 samples per class. In our test results using this data split, the RestNet50 model gave the best classification performance, achieving 96.2% F1-score and 96.2% sensitivity.

We have demonstrated that normal patch-based model training processes with the inclusion of transfer learning are capable of achieving high classification results when applied to gigapixel breast cancer histology images.

7 Conflict of Interest

The authors do not have any personal or financial interests related to the subject matters discussed in this manuscript.

8 Data Availability

The data that support the findings of this study are openly available in "ICNAR-CNR" at https://www.bracs.icar.cnr.it/ and http://doi.org/10.48550/arXiv.2111.04740, reference number [1].

References

[1] Brancati, N., Anniciello, A., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., De Pietro, G., Di Bonito, M., Foncubierta, A., Botti, G. & Others Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database*. **2022** (2022)

[2] Al-Sabaawy, H., Rahawi, A. & Al-Mahmood, S. Standard techniques for formalin-fixed paraffin-embedded tissue: A Pathologist's perspective. *Iraqi J Vet Sci*. **35**, 127-135 (2021)

[3] Reshma, V., Arya, N., Ahmad, S., Wattar, I., Mekala, S., Joshi, S. & Krah, D. Detection of breast cancer using histopathological image classification dataset with deep learning techniques. *BioMed Research International*. **2022** (2022)

[4] Yuan, Z., Huang, H., Jiang, Y. & Li, J. Hybrid deep neural networks for reservoir production prediction. *Journal Of Petroleum Science And Engineering*. **197** pp. 108111 (2021)

[5] Hsin, H. Saliency histogram equalisation and its application to image resizing. *IET Image Processing*. **10**, 787-798 (2016)

[6] Aeffner, F., Zarella, M., Buchbinder, N., Bui, M., Goodman, M., Hartman, D., Lujan, G., Molani, M., Parwani, A., Lillard, K. & Others Introduction to digital image analysis in whole-slide imaging: a white paper from the digital pathology association. *Journal Of Pathology Informatics*. **10**, 9 (2019)

[7] Sirinian, C., Peroukidis, S., Kriegsmann, K., Chaniotis, D., Koutras, A., Kriegsmann, M. & Papanastasiou, A. Cellular Senescence in Normal Mammary Gland and Breast Cancer. Implications for Cancer Therapy. *Genes*. **13**, 994 (2022)

[8] Wu, T., Lei, Z., Lin, B., Li, C., Qu, Y. & Xie, Y. Patch proposal network for fast semantic segmentation of high-resolution images. *Proceedings Of The AAAI Conference On Artificial Intelligence*. **34**, 12402-12409 (2020)

[9] Solem, J. Programming Computer Vision with Python: Tools and algorithms for analyzing images. (" O'Reilly Media, Inc.",2012)

[10] Boschman, J., Farahani, H., Darbandsari, A., Ahmadvand, P., Van Spankeren, A., Farnell, D., Levine, A., Naso, J., Churg, A., Jones, S. & Others The utility of color normalization for AI-based diagnosis of hematoxylin and eosin-stained pathology images. *The Journal Of Pathology*. **256**, 15-24 (2022)

[11] Giaquinto, A., Sung, H., Miller, K., Kramer, J., Newman, L., Minihan, A., Jemal, A. & Siegel, R. Breast cancer statistics, 2022. *CA: A Cancer Journal For Clinicians*. **72**, 524-541 (2022)

[12] Siegel, R., Miller, K., Fedewa, S., Ahnen, D., Meester, R., Barzi, A. & Jemal, A. Colorectal cancer statistics, 2017. *CA: A Cancer Journal For Clinicians*. **67**, 177-193 (2017)

[13] Nawaz, W., Ahmed, S., Tahir, A. & Khan, H. Classification of breast cancer histology images using alexnet. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 869-876 (2018)

[14] Mehra, R. & Others Breast cancer histology images classification: Training from scratch or transfer learning?. *ICT Express*. **4**, 247-254 (2018)

[15] Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., Rao, X., Zheng, C. & Zhang, F. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*. **173** pp. 52-60 (2020)

[16] Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A. & González, F. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. *PloS One*. **13**, e0196828 (2018)

[17] Gour, M., Jain, S. & Sunil Kumar, T. Residual learning based CNN for breast cancer histopathological image classification. *International Journal Of Imaging Systems And Technology*. **30**, 621-635 (2020)

[18] Toğaçar, M., Özkurt, K., Ergen, B. & Cömert, Z. BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics And Its Applications*. **545** pp. 123592 (2020)

[19] Nazeri, K., Aminpour, A. & Ebrahimi, M. Two-stage convolutional neural network for breast cancer histology image classification. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 717-726 (2018)

[20] Brancati, N., Frucci, M. & Riccio, D. Multi-classification of breast cancer histology images by using a fine-tuning strategy. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 771-778 (2018)

[21] Jiang, Y., Chen, L., Zhang, H. & Xiao, X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PloS One*. **14**, e0214587 (2019)

[22] Wei, B., Han, Z., He, X. & Yin, Y. Deep learning model based breast cancer histopathological image classification. *2017 IEEE 2nd International Conference On Cloud Computing And Big Data Analysis (ICCCBDA)*. pp. 348-353 (2017)

[23] Iesmantas, T. & Alzbutas, R. Convolutional capsule network for classification of breast cancer histology images. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 853-860 (2018)

[24] Bejnordi, B., Lin, J., Glass, B., Mullooly, M., Gierach, G., Sherman, M., Karssemeijer, N., Van Der Laak, J. & Beck, A. Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images. *2017 IEEE 14th International Symposium On Biomedical Imaging (ISBI 2017)*. pp. 929-932 (2017)

[25] Wahab, N., Khan, A. & Lee, Y. Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. *Computers In Biology And Medicine*. **85** pp. 86-97 (2017)

[26] Bardou, D., Zhang, K. & Ahmad, S. Classification of breast cancer based on histology images using convolutional neural networks. *Ieee Access*. **6** pp. 24680-24693 (2018)

[27] Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J., Cardoso, F., Siesling, S. & Others Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*. **66** pp. 15-23 (2022)

[28] Gordon, P. The Impact of Dense Breasts on the Stage of Breast Cancer at Diagnosis: A Review and Options for Supplemental Screening. *Current Oncology*. **29**, 3595-3636 (2022)

[29] Belsare, A. & Mushrif, M. Histopathological image analysis using image processing techniques: An overview. *Signal & Image Processing*. **3**, 23 (2012)

[30] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K. & Li, S. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*. **7**, 4172 (2017)

[31] Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A. & Campilho, A. Classification of breast cancer histology images using convolutional neural networks. *PloS One*. **12**, e0177544 (2017)

[32] Joseph, A., Abdullahi, M., Junaidu, S., Ibrahim, H. & Chiroma, H. Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer). *Intelligent Systems With Applications*. **14** pp. 200066 (2022)

[33] Rakhlin, A., Shvets, A., Iglovikov, V. & Kalinin, A. Deep convolutional neural networks for breast cancer histology image analysis. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 737-744 (2018)

[34] Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J., Guan, X., Schmitt, C. & Thomas, N. A method for normalizing histology slides for quantitative analysis. *2009 IEEE International Symposium On Biomedical Imaging: From Nano To Macro*. pp. 1107-1110 (2009)

[35] Pati, P., Jaume, G., Fernandes, L., Foncubierta-Rodriguez, A., Feroce, F., Anniciello, A., Scognamiglio, G., Brancati, N., Riccio, D., Di Bonito, M. & Others Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. *Uncertainty For Safe Utilization Of Machine Learning In Medical Imaging, And Graphs In Biomedical Image Analysis: Second International Workshop, UNSURE 2020, And Third International Workshop, GRAIL 2020, Held In Conjunction With MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*. pp. 208-219 (2020)

[36] Stegmüller, T., Bozorgtabar, B., Spahr, A. & Thiran, J. Scorenet: Learning non-uniform attention and augmentation for transformer-based histopathological image classification. *Proceedings Of The IEEE/CVF Winter Conference On Applications Of Computer Vision*. pp. 6170-6179 (2023)

[37] Cancer.net Breast Cancer (Metastatic) - Statistics. (https://www.cancer.net/cancer-types/breast-cancer-metastatic/statistics), Accessed May 17, 2023

[38] Ott, J., Ullrich, A. & Miller, A. The importance of early symptom recognition in the context of early detection and cancer survival. *European Journal Of Cancer*. **45**, 2743-2748 (2009,9)

[39] Akram, M., Iqbal, M., Daniyal, M., Khan, A., Awareness & Cancer Awareness and current knowledge of breast cancer. *Biological Research*. **50**, 33 (2017,10)

[40] Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M. & Others Bach: Grand challenge on breast cancer histology images. *Medical Image Analysis*. **56** pp. 122-139 (2019)

[41] Vang, Y., Chen, Z. & Xie, X. Deep learning framework for multi-class breast cancer histology image classification. *Image Analysis And Recognition: 15th International Conference, ICIAR 2018, Póvoa De Varzim, Portugal, June 27–29, 2018, Proceedings 15*. pp. 914-922 (2018)

[42] Johnson, J. & Khoshgoftaar, T. Survey on deep learning with class imbalance. *Journal Of Big Data*. **6**, 27 (2019)

[43] Yadav, S. & Jadhav, S. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal Of Big Data*. **6**, 113 (2019)

[44] Hu, B. & Wang, J. Detection of PCB Surface Defects With Improved Faster-RCNN and Feature Pyramid Network. *IEEE Access*. **8** pp. 108335-108345 (2020)

[45] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S. & Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*. **13**, 4712 (2021)

[46] Mehmood, A., Yang, S., Feng, Z., Wang, M., Ahmad, A., Khan, R., Maqsood, M. & Yaqub, M. A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. *Neuroscience*. **460** pp. 43-52 (2021)

[47] Elmore, J., Longton, G., Carney, P., Geller, B., Onega, T., Tosteson, A., Nelson, H., Pepe, M., Allison, K., Schnitt, S. & Others Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*. **313**, 1122-1132 (2015)

[48] Kandel, I. & Castelli, M. How deeply to fine-tune a convolutional neural network: a case study using a histopathology dataset. *Applied Sciences*. **10**, 3359 (2020)

[49] Kour, G. & Saabne, R. Real-time segmentation of on-line handwritten arabic script. *Frontiers In Handwriting Recognition (ICFHR), 2014 14th International Conference On*. pp. 417-422 (2014)

[50] Kour, G. & Saabne, R. Fast classification of handwritten on-line Arabic characters. *Soft Computing And Pattern Recognition (SoCPaR), 2014 6th International Conference Of*. pp. 312-318 (2014)

[51] Hadash, G., Kermany, E., Carmeli, B., Lavi, O., Kour, G. & Jacovi, A. Estimate and Replace: A Novel Approach to Integrating Deep Neural Networks with Existing Applications. *ArXiv Preprint arXiv:1804.09028*. (2018)