Breast Cancer detection using convolutional neural networks and transfer learning

Hee Kyoung Nam - 500462612

I. INTRODUCTION

Breast cancer is the second leading cause of cancer deaths among women overall. According to the World Health Organization (WHO), the number of cancer cases expected in 2025 will be 19.3 million cases. Breast cancer is a type of cancer that starts when cells begin to grow out of control. Thus, early diagnosis and treatments are very important to avoid the progress of the disease.

Invasive Ductal Carcinoma (IDC) is the most common subtype of all breast cancers. During diagnosis, pathologists typically focus on the regions which contain the IDC. However, human-assisted approaches have drawbacks such as expert availability and examination cost. Hence, there is a need for an automatic and accurate histopathological image analysis method, especially classification tasks.

Prior research suggests the potential of an AI-based methodology that incorporates deep learning into pathologist diagnostic systems of cancers such as lung and renal with the hope of minimizing error and improving accuracy [4] [7]. One of the significant advances of diagnosis in pathology is the use of deep neural network for digital image classification. Hence, I propose image classification approaches for breast cancer diagnosis using various convolutional neural network architectures and pre-trained models, Resnet50 and DenseNet201.

II. LITERATURE REVIEW

Several methods have been introduced in the literature to help in detecting breast cancer using whole slide images of the same publicly available IDC histopathological images dataset.

Cruz-Roa et al. in 2014 [1] built a convolutional neural network for IDC automatic tissue classification in breast tissue slide images. Their method yielded the best quantitative results for automatic detection of IDC regions with 84.23% for accuracy and 71.80% for the F-measure score in comparison with a machine learning classifier using random forest.

In another study, Karatayev et al. in 2021 [3] proposed a new CNN model with an accuracy of 92% for the classification of histopathological images which outperforms the baseline model with accuracy of 86%. They used 3 x 3 conv filters and extracted important features using a max pooling operation. Then SoftMax classifier was chosen to output prediction probabilities. In this study, performance of pre-trained models VGG16, ResNet18, and DenseNet achieved 0.76, 0.84, 0.82 for accuracy respectively.

III. METHODS

A. Dataset description

This paper used a publicly accessible image dataset from the Kaggle website. The original dataset consists of 162 whole mount slide images of Breast Cancer specimens scanned at 40x. The dataset includes high-resolution photographs (2040 * 1536 pixels) from 162 positive IDC women at The Cancer Institute of New Jersey. To avoid memory overflow, 100 patient profiles were used instead of 162. From that, 103,351 patches of size 50 x 50 were extracted. There are 78,831 IDC negative samples and 24,520 IDC positive samples. Each patient sample has images of regions without IDC and images of regions with IDC.

B. Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data. By running the analysis, we can discover patterns, spot anomalies, test hypothesis and check assumptions with the help of summary statistics and graphical representations.

I used bar charts to examine the distribution of the number of images available per patient and percentage of IDC containing images per patient. The left chart of Figure 1 shows relatively normal distribution and there are more patients having about 1000 images/patches in the dataset. From the right chart of Figure 1, we see that most of patients have relatively low percentages of IDC positive images in their images. This indicates the dataset is imbalanced in terms of the class labels IDC negative, IDC positive.

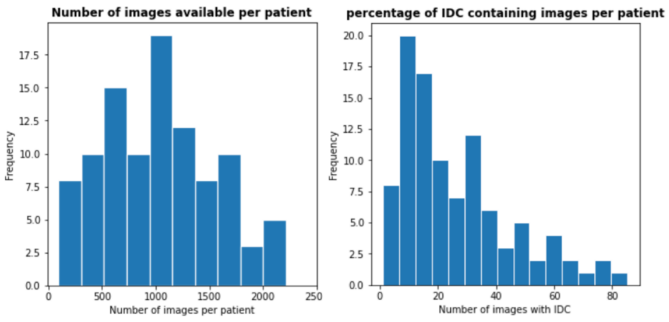

Fig. 1. Number of images available per patient (Left), Percentage of IDC positive images per patient (right)

Before pixel values of images get normalized for training, I have chosen 16 random images from each class IDC absent, IDC present to check if there is any visual difference in images. In figure 3, We can see that IDC absent samples are light colored (light pink) than IDC present images which are closer to violet colors.

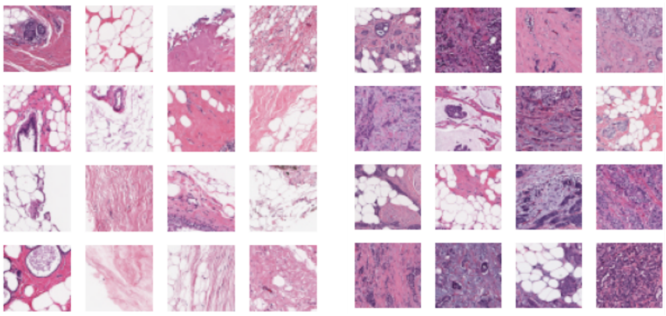

Fig. 2. Random 16 IDC negative samples (left), 16 IDC positive samples (right)

C. Data pre-processing and augmentation

A noticeable problem with this dataset is the unequal distribution of class labels. The size of negative (IDC not present) is about 3 times bigger than the positive (IDC present) data, which harms the deep learning model performance.

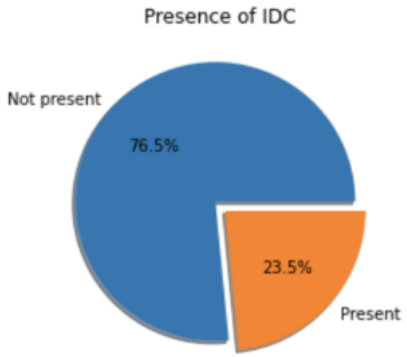

Fig. 3. The ratio between negative and positive data is roughly 75:25 which is not equally distributed

Figure 3 shows that the ratio between negative and positive data is 75:25 and it indicates an imbalanced distribution. In order to solve the imbalance problem, equal number of samples (5000 samples from each class) from both negative and positive data are randomly extracted from the original dataset. Total of 500 samples with 250 negative and 250 positive images were randomly selected for test sets. The sample data are then split into training and test sets with a proportion of 80% and 20% respectively. When splitting the dataset, stratify parameter is used to keep the same proportions of class labels as the input dataset.

Data augmentation is considered as regularization technique that reduce overfitting and improve overall performance of models. The histological images are usually small and challenging to gather. However, Convolutional Neural Networks have been shown to perform well with augmentation in liver injury classification [5] and various medical image issues. Hence, augmentation is often used to overcome issues that come from insufficient training data. In this project, various manipulations such as scaling, rotation, shifting, and flips are applied to the training images. In the process of augmentation, the original training set is further split into train set, having 90% of original training set and validation set, having 10% of original training set for training validation check. Table 1 shows the summary of image augmentation settings.

Augmentation settings	
Augmentation procedure	**Parameter value**
Rescale	1/255
Rotation-range	20
width shift	0.20
height shift	0.20
Horizontal-flip	True
Vertical-flip	True

Table I. Image augmentation settings

D. Model design

A model acts as a function that takes image data as input and returns the class label as output. In this project, three different CNN model architectures and two pre-trained models, ResNet50 and DenseNet201 are examined to test their ability of IDC detection. Architectures in this project are trained on image input size, 50x50x3. The output layers of the architectures have two nodes as there are one hot coded two class labels generated by the image generator and SoftMax activation function is used.

IV. EXPERIMENTAL SETUP

Implementation of this project work are done on google Colab with GPU.

1) CNN architecture: Previous studies have shown that shallow CNNs perform well in IDC detection[1], so I built the first CNN architecture using 3 convolutional layers with 3x3 kernel size, and the layers are followed by max-pooling layers and Relu activation functions. Note that each convolutional layer is designed to have 32, 64, and 128 filters respectively. The second CNN architecture model is formed the same way as the first CNN, but kernel size is changed to 5x5 to examine the effect of kernel size on the testing loss and accuracy. The third CNN architecture is also formed the same way as the first CNN, but batch normalization layer are added to the every convolution layers. The output layer of CNN models has two nodes as the two class labels are one-hot coded so SoftMax activation function is applied.

2) Pre-trained model: Other than CNN approach, I used some of the popular pre-trained models for image classification such as ResNet50 and DenseNet201. ResNet50 is a convolutional neural network with 50 layers and the model was trained on a millions images from ImageNet database [2] for classifying 1000 objects and I will use the weights from this pre-trained model to classify breast cancer classes. The DenseNet is also trained on ImageNet dataset and it has 201 layers which is greater number of layers than the ResNet.

3) Performance measures: Each image patch was classified as either IDC positive or IDC negative with the labels 1 and 0 respectively. The main performance measures used for each model are precision, recall, F1-score, and accuracy values, generated by classification report method. I also compared actual class and predicted results using a confusion matrix.

4) Training parameters: Architectures in this project are trained on 10 epochs. For the model compilation, categorical cross entropy loss function is selected, and Adam optimization technique is used with learning rate 1e-4. Adam is known to be the best optimizer for image classification and previous studies have used Adam optimizer for IDC detection[6].

Hyperparameter	value
Batch size	10
Number of epochs	10
Optimizer	Adam
Initial learning rate	1e-4

Table 2. Training hyperparameters

V. RESULTS

A. CNN model

The first CNN model with three convolutional layers followed by max pooling layer and kernel size (3x3) resulted in an overall accuracy of 85% for the test set. The second CNN model with kernel size (5x5) resulted in an overall accuracy value of 83%. The third CNN model with three convolutional layers followed by batch normalization and max pooling layers resulted in an overall accuracy value of 83%.

The differences in overall accuracy value between three CNN architectures are very mild. Since there is not much noticeable observation made between CNN models, I will present results of the first CNN model only. The first CNN model correctly predicted 424 out of 500 test instances. 202 out of 250 are true IDC negative and 48 are actually IDC negative. 222 of 250 are true IDC positive and 28 IDC positive instances are misclassified as IDC negative (Figure 4).

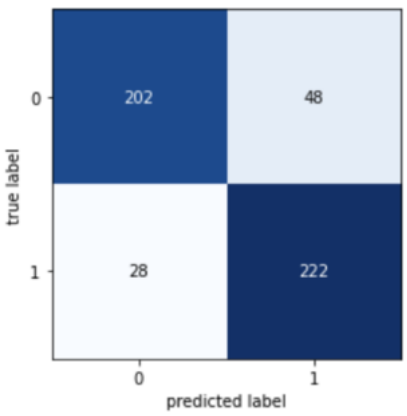

Fig. 4. Confusion matrix of the first CNN model {0: IDC negative, 1: IDC positive}

	Precision	Recall	F1-Score	Support
IDC negative	0.88	0.81	0.84	250
IDC positive	0.82	0.89	0.85	250
macro avg	0.85	0.85	0.85	500
weighted avg	0.85	0.85	0.85	500

Table 3. Performance measures of the customized CNN on testing set.

B. ResNet50

The Resnet50 model correctly predicted 326 out of 500 test instances. 107 out of 250 are true IDC negative and 143 are actually IDC positive, and 219 were correctly predicted as IDC positive, while 31 IDC negative instances were misclassified (Figure 5). ResNet50 model achieved an overall accuracy of 65% on test set .

	Precision	Recall	F1-Score	Support
IDC negative	0.78	0.43	0.55	250
IDC positive	0.60	0.88	0.72	250
macro avg	0.69	0.65	0.63	500
weighted avg	0.69	0.65	0.63	500

Table 4. Performance measures of the ResNet on testing set.

C. DenseNet201

The Densenet210 model correctly predicted 418 out 500 test instances. 191 out of 250 are true IDC negative and 59 are actually IDC positive. 227 out of 250 are true IDC positive and 23 IDC negative instances are misclassified (Figure 5). Densenet201 model achieved an overall accuracy of 84% on test set.

	Precision	Recall	F1-Score	Support
IDC negative	0.89	0.76	0.82	250
IDC positive	0.79	0.91	0.85	250
macro avg	0.84	0.84	0.84	500
weighted avg	0.84	0.84	0.84	500

Table 5. Performance measures of the DenseNet on testing set.

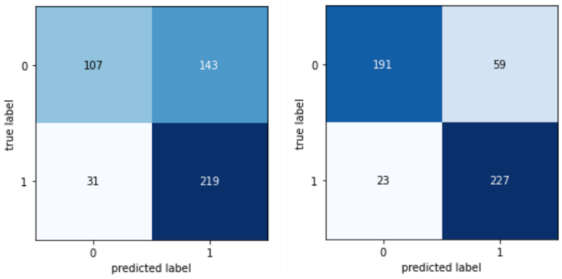

Fig. 5. Confusion matrix of ResNet50 (left) and DenseNet201 (right) {0: IDC negative, 1: IDC positive}

Model	Loss	Accuracy
CNN (1)	0.40	85%
ResNet	0.59	65%
DenseNet	0.40	84%

Table 6. Results of first customized CNN, ResNet50, DenseNet201 model on IDC dataset

VI. CONCLUSION

This project proposed convolutional neural network architectures, ResNet and DenseNet pre-trained models for breast invasive ductal carcinoma classification from histopathological images. The proposed CNN models are customized and three different architectures are formed to examine the effect of kernel size and batch normalization layers. However, no significant difference was observed between them. I assume that higher number of epoch size might have changed the results. In this project due to memory shortage, epoch size of 10 was selected which is a relatively low epoch size.

Two pre-trained models, ResNet50 and DenseNet201 are tested on IDC dataset. ResNet did not perform well in detecting IDC regions of breast tissue slides but DenseNet performed well in detection. Studies have shown that ResNet can detect IDC regions of breast tissue slides [3]. I assume that I had low performance of ResNet in this project because of relatively small sample size. A study done by Karatayev et al. [3] tested ResNet on 100,000 IDC images which is 10 times more samples than the samples used for this project. Small epoch size and sample size due to the low memory environment are parts of the limitations I experienced in this project.

Through this project, I learned to augment image data for overfitting prevention and using pre-trained model is powerful approach. Pre-trained model required less training and effort in building the classification model. Whereas shallow CNN model worked pretty well in IDC detection but it was challenging to customize suitable CNN model with different configuration/layer options.

For the future work, we can expand this work to evaluate other CNN models and extend the model usability to other types of cancer image dataset. I hypothesize that a composite model that utilizes more than one architecture may be able to produce better classification performance than any single architecture. Thus, Further research may focus on developing a multi-model deep learning approach as well.

REFERENCES

[1] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N.N.C. Shih, J. Tomaszewski, F.A. González, A. Madabhushi, Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent, Sci. Rep. (2017) 7, doi:10.1038/srep46450.

[2] ImageNet. http://www.image-net.org

[3] M. Karatayev, S. Khalyk, S. Adai, M. H. Lee, and M. F. Demirci, “Breast cancer histopathology image classification using CNN,” 2021 16th International Conference on Electronics Computer and Computation (ICECCO), 2021. doi: 10.1109/ICECCO53203.2021.9663757.

[4] Ozdemir O, Russell RL, Berlin AA. A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans. IEEE Trans Med Imaging (2019) 39(5):1419–29. doi: 10.1109/TMI.2019.2947595

[5] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning", Journal of Big Data, vol. 6, no. 1, pp. 1-48, 2019.

[6] M. A.-E. Zeid, K. El-Bahnasy, and S. E. Abo-Youssef, “Enhanced CNN architecture for invasive ductal carcinoma detection in breast histopathology images,” 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS), 2021.

[7] Xi IL, Zhao Y, Wang R, Chang M, Purkayastha S, Chang K, et al. Deep Learning to Distinguish Benign from Malignant Renal Lesions Based on Routine MR Imaging. Clin Cancer Res (2020) 26:1944–52. doi: 10.1158/1078-0432.CCR-19-0374