Breast cancer histopathology image classification using CNN

Mukhamejan Karatayev
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
mukhamedjan.karatayev@nu.edu.kz

Saltanat Khalyk
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
saltanat.khalyk@nu.edu.kz

Min-Ho Lee
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
minho.lee@nu.edu.kz

Shomanov Adai
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
adai.shomanov@nu.edu.kz

M Fatih Demirci
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
muhammed.demirci@nu.edu.kz

***Abstract*—The breast cancer is one of the wide spread diseases around the world. Cancer develops in a milk duct and then spreads to the surrounding breast tissues. This initial stage of progression is called invasive ductal carcinomas (IDC). Almost 80% of all breast cancers are invasive ductal carcinomas. If IDC is detected at early stages, the patient can be treated and will have a high survival rate, whereas undetected cancer may spread into other parts of the body, as well as surrounding breast tissues. In this work, the dataset that contains breast cancer histopathology images was used. The objective of this work is to implement a convolutional neural network (CNN) model for accurate IDC classification, by balancing the dataset and tuning hyperparameters. The proposed model achieves an accuracy of 92% for the classification of histopathological images, and outperforms the baseline CancerNet model with accuracy of 86%. Furthermore, our experimental results demonstrate the superiority of our approach over the pre-trained networks, such as VGG16, DenseNet and ResNet18.**

***Index Terms*—Deep Learning, CNN, breast cancer, IDC, CancerNet.**

I. Introduction

Nowadays, mankind is faced with various kinds of diseases. One of the most common and dangerous is breast cancer. According to the World Health Organization, 2.3 million women were diagnosed with breast cancer and there were 685 000 deaths globally in 2020. Almost 7.8 million women were diagnosed with breast cancer who still have a progressive disease at the end of 2020 [4]. IDC is Invasive Ductal Carcinoma; cancer that develops in a milk duct and invades the fibrous or fatty breast tissue outside the duct; it is the most common form of breast cancer forming 80% of all breast cancer diagnoses. Histology is the study of the microscopic structure of tissues.

Motivation for this work is to help to identify breast cancer as soon as possible and reduce human error, because the early detection of cancer can help a person to defeat the disease or increase the person's lifetime. There are many methods which are used for cancer detection, but in this work we will use histopathological images which reveal cancer cells in a human tissue. Those images present the difference between cancer and normal cells, which makes it possible to build an image classification model to detect cancer cells.

There are many interesting works and experiments regarding breast cancer classification. The most recent studies are showing 96.5% accuracy with Random Forest and Support Vector Machine [1]. However, the authors of this work used the Wisconsin Breast Cancer Dataset from the 1995 year and dataset size is relatively small, including 357 benign and 212 malignant cases. In comparison, our dataset has 277,524 images.

This same Wisconsin Breast Cancer dataset was trained with CNN of 3 layers and having rectified linear units between hidden layers, and had F1 score of 98 for the benign class and F1 score of 99 for the malignant class [9].

Another interesting research was done using digital mammograms datasets from MIAS and classified them with a self-regulated multilayer perceptron neural network. The dataset proportions were as follows: 56 benign, 39 malignant and 75 normal cases. The accuracy of predictions are 90% [2].

In the works of Andrew Janowczyk, invasive ductal carcinoma detection was done with 50k testing patches and reached an F-score of 0.7648 using deep learning models [3].

Irum Hirra, Mubashir Ahmad cropped the histopathological images to extract the region of interest, then converted the RGB images to grayscale images. They did thresholding to get rid of low intensity images. After that, patches were created which have two regions: background region and main region. If the main region of interest in the given patch is more than 70%, then the image is classified as having cancer. Once every image was preprocessed, they were trained on a novel patch-based deep learning model. The four datasets with 3002×2384 image resolution from Hospital of the University of Pennsylvania (HUP), Case Western Reserve University (CWRU), Cancer Institute of New Jersey (CINJ), and The Cancer Genome Atlas (TCGA) were used. This proposed network achieved 86% accuracy [18].

It is interesting to note that the breast cancer classification is a huge topic of interest and great research involving image analysis using deep learning techniques was done not only with histopathology images but also with Magnetic Resonance Imaging (MRI), mammogram X-ray images.

Nur Syahmi Ismail, Cheab Sovuthy did breast mammogram images classification using IRMA dataset using deep learning models such as ResNet50, VGG16, and reached accuracy of 91.7% and 94%, respectively [15]. The original image size was 128×128 and was resized to 224×224 to fit the network architecture. The benign image amount was 931 and malignant was 538. In comparison, we could not resize our histopathology images to 224×224, because of the big difference in image resolution.

Also, nowadays there are many object detection algorithms which help to increase the prediction accuracy of deep learning models. In particular, You Only Look Once (YOLO) and RetinaNet were used to detect breast cancer from mammogram images in the works of Hamed et.al [16].

Image filtering techniques for removing noise, such as median filter, contrast-limited adaptive histogram equalization, were applied in the works of Hao-Chun Lu, El-Wui Loh. The F1 score was 0.79, 0, 0.88, whereas accuracy reached 82%. The X-ray mammogram images had more benign cases that's why the data augmentation was used to balance the dataset. Overall, they had trained their CNN with almost 10 000 images from Taiwan hospital with 2294×2294 image resolution [17].

In this work, we will implement CNN architecture for the IDC histopathological image classification. Then, it will be compared with existing CNN methods and CancerNet network. A set of comprehensive experiments on the IDC dataset shows that our approach outperforms all proposed methods discussed previously.

The remainder of this paper is organized as follows: Section II introduces methodology of this work. Section III reports our experiments and describes our results. In Section IV we will discuss the issues and propose suggestions for the future work. Finally, Section V concludes the paper outlining key achievements of this project.

TABLE I
IMBALANCED DATASET USED FOR TRAINING RESNET18, DENSENET

Dataset	IDC	non IDC
train	63,028	158,990
test	7,880	19,875
validation	7,878	19,873

TABLE II
BALANCED DATASET FOR PROPOSED CNN

Dataset	IDC	non IDC
train	45,000	55,000
test	10,000	15,000

II. METHODOLOGY

Deep learning (DL) explores the possibility of learning features directly from input data, avoiding hand-crafted features [10]. The main concept of DL is to discover multiple levels of representation aiming that higher-level features represent more abstract semantics of the data. The Convolutional Neural Networks (CNNs) technique has achieved great success in image classification problems, including medical image analysis [10], [13], [14]. Therefore, we proposed multiple-layer CNN network and compared it with torchvision image classification models, such as: VGG16, DenseNet, ResNet18. The main focus was on the dataset preparation, as it has significant impact on the accuracy of the model.

A. Dataset

We used the IDC regular dataset (the breast cancer histology image dataset) from Kaggle [11]. This dataset holds 277,524 patches of size 50×50 extracted from 162 whole mount slide images of breast cancer specimens scanned at 40x. Of these, 198,738 test negative and 78,786 test positive with IDC. Figure 1 represents the random sample of images from the dataset. It is important to note that the original dataset was split by patient id and each of them had IDC and non IDC images. For the comparison, we trained the ResNet18 and DenseNet with an imbalanced dataset to preserve the large database of IDc images. The dataset was split to train, test, validation and their proportions can be seen in Table I.

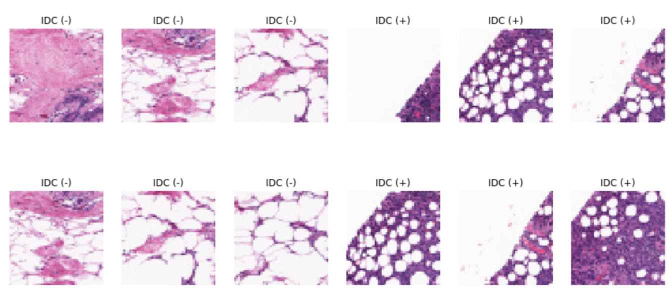

Fig. 1. Sample images from IDC dataset

As we can see, the initial dataset in Table I is highly imbalanced with 71.6% negative and only 28.3% positive samples. Therefore, this dataset was split into train and test sets with balanced class values. As a result, Table II shows the distribution of classes in train and test sets, which was used to train proposed CNN model.

B. Models

In this work, we implemented CancerNet as a baseline model [5], then trained VGG16, ResNet18, DenseNet using pytorch pretrained models, and proposed a new CNN model built on Keras network.

The baseline model is the Convolutional Neural Network (CancerNet) trained on the image dataset [5]. It uses 3×3 Conv filters, stacking these filters on top of each other. Then to extract the important features, a max-pooling operation was performed. Also, they used depthwise separable convolution which is more efficient and takes up less memory. For that Sequential API was used to build CancerNet and Separable-Conv2D. The network predicts one of the two classes, where 0 is non-IDC and 1 is IDC. There are three [DEPTHWISE CONV = RELU = POOL] layers, each having a higher stacking and a greater number of filters. Finally, the softmax classifier outputs prediction percentages for each class. Data augmentation was used to generalize the model.

The ResNet18 is a convolutional neural network with 18 layers [6]. Original ResNet18 was trained on more than 14 millions of images from ImageNet database for classifying 1000 objects, and we will use its weights to classify our dataset to malignant or benign breast cancer classes. Therefore, we have used pretrained ResNet18 from torch models and changed the last layer to output only 2 classes. The network was trained on image input size of 224×224, therefore we also changed the input size to 50×50.

The DenseNet is similar to ResNet18, but as the name says, it has more dense layers. ResNet18 gets only one input from the preceding layer and passes it forward to the next layer. In contrast, DenseNet uses the output or the feature maps from previous layers to feed the every next layer as inputs. For example, if there are 5 layers, then the first feature map extracted from first layer, will be used to feed each of the 5 layers. Finally, the input was resized to 50×50 RGB images.

The VGG16 is a convolutional neural network with 16 layers, also trained on ImageNet dataset. In this case, we passed the 50×50 sized RGB image to convolutional layers, then 3×3 filters were used, followed with max-pooling. Fully connected layers have different depths and the last layer performs a softmax activation function. The rectified linear activation function (ReLu) was used between hidden layers.

Figure 2 represents CNN architecture that provided the best results in our experiments. Overall, we have four (3×3) CONV2D layers each with a higher stacking and a greater number of filters, followed by max-pooling layers and Relu activation functions. The last one, a fully-connected output layer with softmax activation, depends on the number of classes in the classification problem, i.e., 2 output filters for our binary classification problem.

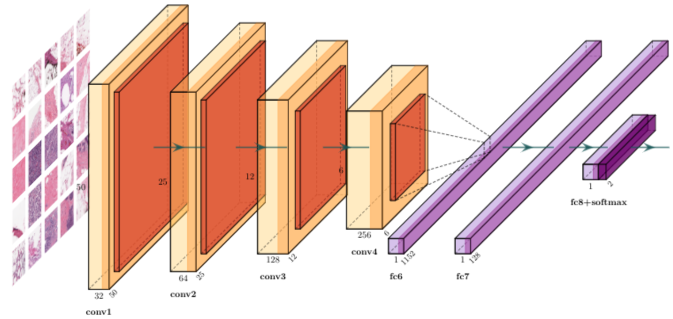

Fig. 2. Proposed CNN architecture

C. Training

In total, four CNN models were trained in this study using the data described in Section 2, whereas baseline CancerNet was used as comparison. The trained models include VGG16, ResNet18, DenseNet, and our CNN network. The CNN models were trained on a NVIDIA GTX 2070 GPU using the PyTorch framework. Training took about 60 minutes for each model. The Stochastic gradient descent (SGD) was used as an

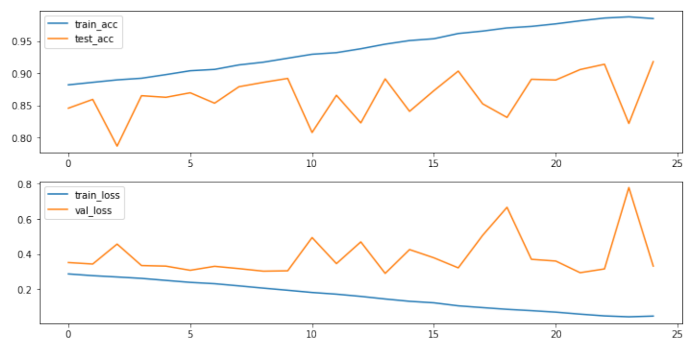

Fig. 3. Proposed CNN model training loss and accuracy

optimizer with Binary cross-entropy loss function. Conducting multiple experiments, it was found that 25 epochs is the most optimal value, as the accuracy of the model did not increase after this limit. Also, the learning rate of 0.001 and momentum of 0.9 allowed the model to learn a more optimal or even globally optimal set of weights. Figure 3 demonstrates the training loss and test accuracy of the CNN model introduced in this paper. The performance of the model was evaluated by using accuracy, precision, recall and F1-score.

III. Results

All five models were tested on the same test set. The experimental results have shown that our CNN network outperforms all other approaches in breast cancer image classification problem. According to the results, VGG16 obtained the worst performance with 76%, next is the DenseNet with 82%, then ResNet18 with 84%, and finally our CNN model with 92% accuracy on a test set. The overall test results are shown in

Table III and the confusion matrix for the proposed CNN model can be seen in Table IV.

A. Accuracy

TABLE III
RESULTS OF VGG16, RESNET18, DENSENET, CANCERNET AND PROPOSED CNN ON IDC DATASET

Model	Epoch	Learning rate	Loss	Accuracy
VGG16	25	0.01	1.94	0.76
ResNet18	25	0.01	0.40	0.84
DenseNet	25	0.01	0.45	0.82
CancerNet	40	0.01	0.38	0.86
Proposed CNN	25	0.001	0.33	0.92

The best performance was obtained with the proposed CNN model and its confusion matrix in Table IV.

TABLE IV
RESULTS OF OUR CNN ON KERAS

Classes	precision	recall	f1-score	support
0	0.95	0.89	0.92	10000
1	0.89	0.95	0.92	10000
accuracy			0.92	20000
macro avg	0.92	0.92	0.92	20000
weighted avg	0.92	0.92	0.92	20000

B. Visualization

It was found that patches with positive breast cancer look more violet and crowded compared to the negative ones. Thus, there are some differences which can be learned by neural networks to separate unhealthy patients from others. Also, the shape of the patches and their distribution in each mammogram was analyzed using Binary objective visualization for each tissue slice. According to Figure 4, there is a large variation in the concentration of cells between different samples.

IV. DISCUSSION

According to the results, the DL approaches have proven to be effective for breast cancer image classification problems. They were able to successfully extract required features of cancer tissues from the images. The underlying architecture of all tested CNN networks was similar to each other. Therefore, the proper data preparation and hyperparameters tuning are crucial for improvement of classification accuracy.

As we said before, balancing the dataset significantly improved the performance of our CNN model. Also, as a result of experiments we were able to find the most optimal hyperparameters. We tested different learning rates, optimizers, number of epochs and CNN architectures. The presented network is the best solution found during this research.

For future work, it would be interesting to test the model on a completely different dataset, to test its robustness. Also, IDC dataset can be augmented using different methods, which may improve the performance even further.

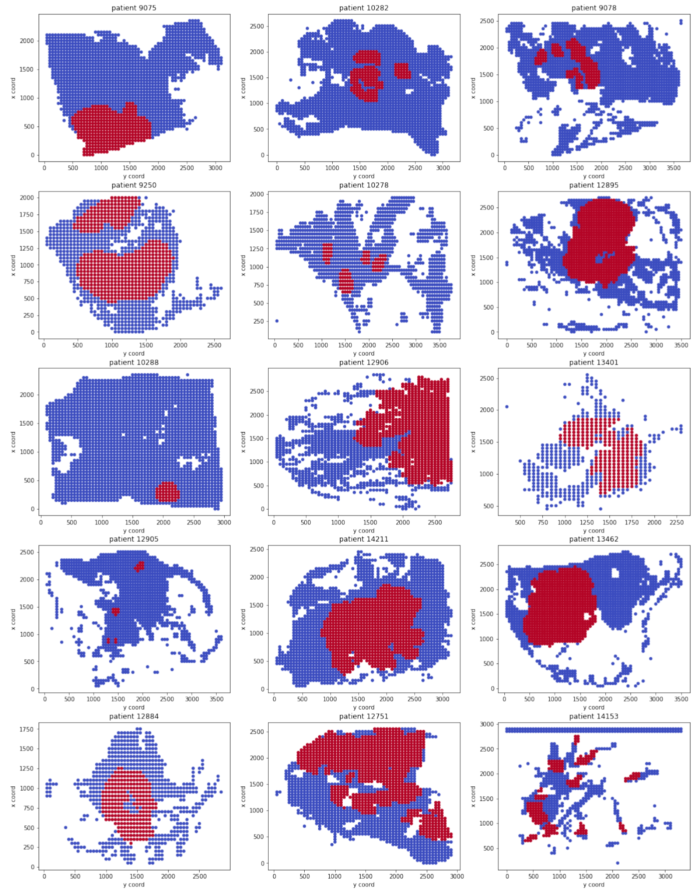

Fig. 4. Binary objective visualization for each tissue slice

V. CONCLUSION

In the presented work we implemented a CNN model for breast cancer recognition based on the IDC dataset (with histopathological images of invasive ductal carcinoma). Initially, we did a literature review of relevant work among the most common subtypes of all breast cancers, as well as available datasets. As a result, we chose the CancerNet network as the baseline approach, also trained multiple CNN networks from PyTorch library, to compare them with our solution. We were able to obtain acceptable results with all CNN models, however the presented CNN architecture was superior. One of the main reasons for such results is the properly balanced dataset used for training and testing. To conclude, we discovered that modern CNN models can be effectively used to detect breast cancer tissues from histopathology scan images, by this saving someone's life.

REFERENCES

[1] Ara, S., Das, A., & Dey, A. (2021, April). Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. In 2021 International Conference on Artificial Intelligence (ICAI) (pp. 97-101). IEEE.

[2] F. F. Ting and K. S. Sim, ”Self-regulated multilayer perceptron neural network for breast cancer classification,” 2017 International Conference on Robotics, Automation and Sciences (ICORAS), 2017, pp. 1-5, doi: 10.1109/ICORAS.2017.8308074.

[3] Janowczyk, Andrew, and Anant Madabhushi. “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases.” Journal of pathology informatics vol. 7 29. 26 Jul. 2016, doi:10.4103/2153-3539.186902

[4] Breast cancer. (n.d.). Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[5] DataFlair (2021). Baseline source code. Retrieved from https://data-flair.training/blogs/project-in-python-breast-cancer-classification/

[6] Resnet18. Retrieved from www.mathworks.com/help/deeplearning/ref /resnet18.html

[7] DenseNet Architecture Explained with PyTorch Implementation from TorchVision. Retrieved from https://amaarora.github.io/2020/08/02/densenets.html

[8] 1. N. S. Ismail and C. Sovuthy, ”Breast Cancer Detection Based on Deep Learning Technique,” 2019 International UNIMAS STEM 12th Engineering Conference (EnCon), 2019, pp. 89-92, doi: 10.1109/EnCon.2019.8861256. (dataset: IRMA 2008)

[9] S. S. Prakash and K. Visakha, ”Breast Cancer Malignancy Prediction Using Deep Learning Neural Networks,” 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 88-92, doi: 10.1109/ICIRCA48905.2020.9183378. (dataset - Wisconsin breast cancer UCI 1992)

[10] N. Khuriwal and N. Mishra, ”Breast Cancer Detection From Histopathological Images Using Deep Learning,” 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), 2018, pp. 1-4, doi: 10.1109/ICRAIE.2018.8710426. (dataset- MIAS Mammography)

[11] Kaggle Inc. (2021). Dataset: IDC regular dataset (the breast cancer histology image dataset) from Kaggle. Retrieved from https://www.kaggle.com/paultimothymooney/breast-histopathology-images/

[12] DataFlair (2021). CNN source code. Retrieved from https://www.kaggle.com/chsharshavardhan/keras-acc-91

[13] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016, July). Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) (pp. 2560-2567). IEEE.

[14] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. BioRxiv, 242818.

[15] N. S. Ismail and C. Sovuthy, ”Breast Cancer Detection Based on Deep Learning Technique,” 2019 International UNIMAS STEM 12th Engineering Conference (EnCon), 2019, pp. 89-92, doi: 10.1109/EnCon.2019.8861256.

[16] Hamed, G., Marey, M. A. E. R., Amin, S. E. S., & Tolba, M. F. (2020, April). Deep learning in breast cancer detection and classification. In The International Conference on Artificial Intelligence and Computer Vision (pp. 322-333). Springer, Cham.

[17] Lu, H. C., Loh, E. W., & Huang, S. C. (2019, May). The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection. In 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 9-12). IEEE.

[18] I. Hirra et al., ”Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling,” in IEEE Access, vol. 9, pp. 24273-24287, 2021, doi: 10.1109/ACCESS.2021.3056516.