Comparison of Cutting Edge Convolutional Neural Network for Breast Cancer Histopathology Image Diagnosis

Tongyuan Qian*

Virginia Polytechnic Institute and State University, N Main St, Blacksburg, USA
**tongyuanqian@vt.edu*

Abstract:
Female breast cancer, especially the invasive ductal carcinoma, is a very common type of cancer. When breast cancer is diagnosed and treated at an early stage, patients can achieve a higher survival rate. In recent years, neural networks have shown their potential in medical fields. Therefore, this work focused on applying three state-of-the-art convolutional neural networks (CNNs), namely ResNet50V2, InceptionV3 and VGG16 to diagnosing breast cancer from histopathology images to verify whether CNNs can be an effective tool in this case. The three architectures were trained with an original dataset of breast cancer histopathology images. After the image pre-processing and the hyperparameter tuning, the evaluation and comparison of the networks' performance were performed. They were evaluated through several statistical analysis based on accuracy, recall, precision, F1-score and training time. The experimental results showed that InceptionV3 obtained the best performance with the accuracy of 87.12% and F1-score of 86.99. ResNet50V2 achieved a close performance with a 74% training time compared with InceptionV3. The result proved that the state-of-the-art CNNs can be considered as a supportive tool that can help diagnosing breast cancers.

Keywords: *Breast cancer detection; Convolutional Neural Network; ResNet50V2; Inception V3; VGG 16*

1 INTRODUCTION

According to recent estimates from the International Agency for Research on Cancer (IARC), female breast cancer is the most common cancer. About 2.3 million women were diagnosed with breast cancer and 685, 000 deaths were caused by breast cancer in 2020 [11]. Despite this, breast cancer typically has a good prognosis with timely diagnosis and proper treatment. The number of women alive today and diagnosed within five years is estimated to be around 8 million. This is higher than the survivor number for any other cancer type [11]. The overall 5-year relative survival rate is 90%. For the localized stage, it is 99%. The survival rate decreases to 86% for the regional stage and 29% for the distant stage [1]. Among all types of breast cancer, invasive ductal carcinoma (IDC) is the most common type and accounts for more than 80% of invasive breast cancers [2]. Diagnosing IDCs can be time-consuming and complex. Pathologists need to look carefully at the magnified histopathology pictures to judge. Therefore, it is essential to develop tools to help pathologists diagnose more efficiently and reduce errors.

Currently, histopathology images have become simpler to obtain, and computer graphics and deep learning techniques have developed a lot. Increasingly deep learning techniques, especially convolutional neural networks (CNNs), are being applied to the medical field to diagnose diseases [4]. In another study [3], the DenseNet201 CNN architecture obtained 96.29% accuracy at diagnosis of COVID-19. In addition, the researcher [7] implemented a CNN model that could reach 97.37% accuracy on malaria diagnosis. In addition, a large amount of research shows that CNNs can be applied as an aid in the image-based diagnosis of different diseases. Breast cancer images have a complex geometrical shape that makes them challenging to diagnose manually. Therefore, many studies have developed various CNNs based on different breast cancer datasets. In recent study of this topic, the CNN model achieved 85.41% balance accuracy on diagnosing IDC from histopathology images [8]. However, most studies did not address and compared those pre-established cutting-edge CNN architectures so that it is hard to find which CNN model can be more suitable in the breast cancer dataset

In this paper, the performance of several developed CNN architectures was evaluated to determine which one is more suitable for diagnosing the IDC. The ResNet50V2, VGG16, and InceptionV3 architectures were compared in terms of accuracy, recall, precision, F1 score and training time. This paper will start with the introduction of the dataset and the image pre-processing. Then, architectural details about the general CNN were explained, ResNet50V2, VGG16, and InceptionV3 and how they can be used in this problem. All in all, this paper will discuss the result and performance of each CNN architecture with the specs mentioned above.

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 664-669, 2023.
https://doi.org/10.2991/978-94-6463-040-4_101

2 METHOD

2.1 Dataset Description

The original dataset of this paper is from Kaggle's Breast Histopathology Images dataset [6]. The images were cropped from another dataset that consisted of 162 whole mount slide images of breast cancer specimens scanned at 40x. Figure 1 is an example specimen image. The dataset contains 277, 524 RGB images with 50 × 50 pixels. There are 198, 738 IDC negative and 78, 786 IDC positive samples. Each image file's name was labelled in this format: <ID>_x<X>_y<Y>_class<C>.png. Where ID is the patient ID, X is the x-coordinate of the original specimen image, Y is the y-coordinate of the original specimen image and C is the sample class where 0 is negative, and 1 is positive. Figures 1 and 2 show some examples of the dataset and naming format.

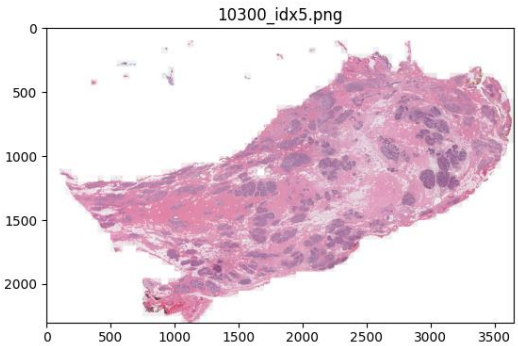

Figure 1: Tissue image reassembled from the dataset with patient ID 10300.

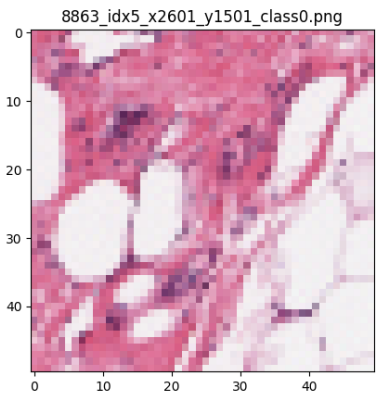

Figure 2: A negative sample with patient ID 8863.

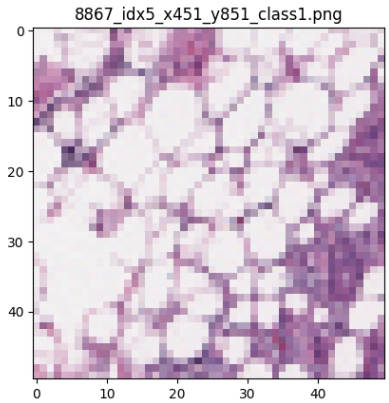

Figure 3: A positive sample with patient ID 8867.

2.2 Data Pre-processing

Since this is an unbalanced dataset and because of computer performance limitations, in this paper, 10,000 samples were taken randomly from each of the positive and negative data. Some of the samples with dimensions not equal to 50 × 50 pixels were replaced with other samples. A total dataset of 20,000 was randomly shuffled and divided into a training and validation set at a ratio of 80% and 20%. This dataset was normalized differently in this study. For ResNet50V2, the sample images were resized to 224 × 224 pixels with RGB values scaled between -1 and 1. For VGG16, the sample images were resized to 224 × 224 pixels with RGB values converted to BGR. For InceptionV3, the sample images were resized to 299 × 299 pixels with RGB values scaled between -1 and 1. No changes were made to the labels of each sample during this processing.

2.3 Convolutional Neural Network

The convolutional neural network (CNN) is a type of deep learning neural network architecture. The powerful image recognition and image processing capabilities are derived from its three main build blocks: convolution layers, pooling layers, and fully connected layers. It is also designed to automatically and adaptively learn to extract essential features through the backpropagation algorithm.

A convolutional layer applies the convolution operations and activation functions to realize the feature extraction. This layer uses kernels to perform element-wise product with part of the input tensor and sum up the product result to compose the feature map. The kernel's size and number can be adjusted with different input and output requirements. The weights in the kernel can be adjusted through the backpropagation algorithm to extract more essential features. The feature map is then passed through the nonlinear activation function, such as the rectified linear unit (ReLU), to get the outputs of this layer.

A pooling layer performs a downsampling operation. This layer can reduce the size of the feature map to decrease the number of parameters and obtain the translation invariance. A max-pooling layer selects the maximum element from the region covered by the filter, whereas an average-pooling layer calculates the average of the elements present in the region.

A fully connected layer flattens the output feature maps of the last convolutional layer or pooling layer and is connected to all the outputs with a weight and an activation function like ReLU. The final layer would use an activation function such as softmax to normalize the output values to class probabilities. The node with the highest probability will become the predicted class.

2.3.1 ResNet50V2

ResNet50V2 is a very deep network with 50 layers. Unlike normal CNNs, ResNet inserts shortcut connections that turn the network. These shortcuts help reduce the vanishing gradients and the degradation problem. As a result, the efficiency of training is increased, and deeper networks can still achieve sufficient improvement. This net achieved 96.5% top-5 test accuracy in ImageNet [5].

2.3.2 VGG16

VGG16 is a CNN model with 16 weighted layers with 13 convolutional layers and 3 dense layers. Five max-pooling layers are also used. It uses small 3 x 3 receptive kernels throughout the whole neural network. The first two convolution layers have 64 kernels, and the number of kernels increases layer by layer until it reaches 512. The VGG16 model achieved 92.7% top-5 test accuracy in ImageNet [9].

2.3.3 InceptionV3

InceptionV3 uses inception modules as basic architectural blocks to compose the whole structure. The inception module consists of convolution kernels of sizes from 1 × 1 to 5 × 5. This feature lets the network identify image features at different sizes and reduces parameter numbers. The model reached 94.49% top-5 test accuracy in ImageNet [10].

Table 1: Summary of architectures.

Network	Depth	Parameters (Millions)
ResNet50V2	50	25.6
VGG16	16	134.3
InceptionV3	48	21.8

2.4 Training Methodology

The three CNNs (i.e. ResNet50V2, VGG16 and Inception V3) were trained with the same hyperparameters. They were trained with batch size 32 for 50 epochs. Stochastic gradient descent was used as the optimizer with an initial learning rate of 0.01. If there were a learning plateau in validation accuracy, the learning rate would be decreased by a factor of 0.1. Since there are only two categories, positive and negative, the final layer of each model was connected with a sigmoid activation function, and binary cross-entropy was used as the loss function.

2.5 Evaluation

The performance of the architectures for this research was evaluated with several matrices: accuracy, recall, precision, and F1-score.

Accuracy measures the percentage of a binary classification test correctly identifies a condition. It is represented in Equation (1), where TP represents the true positives, TN represents the true negatives, FP represents the false positives, and FN represents the false negatives.

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

Recall measures the percentage of the samples in positive class that classified correctly, which can be formulated in Equation (2).

$$Recall = \frac{TN}{TP + FN} \quad (2)$$

Precision measures the percentage of actual positives in predicted positives, which can be formulated in Equation (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

F1-score that represented by Equation (4) is the harmonic mean of recall and precision. This value generally shows represents a model's performance on positive classes.

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

3 RESULT AND DISCUSSION

3.1 Performance for CNNs

In this research, three CNN architectures were applied to predicting IDC with images. The accuracy, recall, precision, and F1-score were recorded to be evaluated during the training process. The time is measured in seconds per epoch. The Table 2 shows the results.

Table 2: Evaluation results of architectures.

Measures	ResNet50V2	InceptionV3	VGG16
Accuracy	86.78	87.12	84.10
Recall	86.43	88.53	85.35
Precision	86.38	85.50	82.55
F1-score	86.40	86.99	83.93
Time (s)	43	58	58

The results showed that ResNet50V2 and InceptionV3 had relatively similar performance, while the VGG16 had the worst performance in this case. Starting with the accuracy section, InceptionV3 achieved the best accuracy of 87.12%, followed closely by ResNet50V2 of 86.78%. VGG16's accuracy was 84.10%. In the recall metric, VGG16 obtained the lowest result with 85.35%, while ResNet50V2 got 86.43% and the highest score was achieved by the InceptionV3 with 88.53%. Differently, under the precision section, ResNet50V2 got the best score of 86.38%, which followed with InceptionV3's 85.50%. VGG16 had 82.55% precision. Based on the above recall and precision results, InceptionV3 obtained the highest F1-score of 86.99%, ResNet50V2 got the secondary of 86.40% and VGG16 got the least score of 83.93%. In addition, the training time of the ResNet50V2 was less than others. Compared with InceptionV3, ResNet50V2 achieved a similar performance with 74% amount of processing time. Table 3 presents the confusion matrix of the InceptionV3's validation test. 3485 out of 4000 samples was predicted correctly.

Table 3: Confusion matrix based on InceptionV3.

Total 4000	Predicted Positive	Predicted Negative
Actual Positive	1722	223
Actual Negative	292	1763

Furthermore, in the study, all three neural networks were found to have different degrees of overfitting. For example, As shown in Figure 4 and 5, the InceptionV3 model started overfitting at around epoch 15 and the final gap between the accuracy and loss at the end of epoch 50 was 4.63% and 0.187.

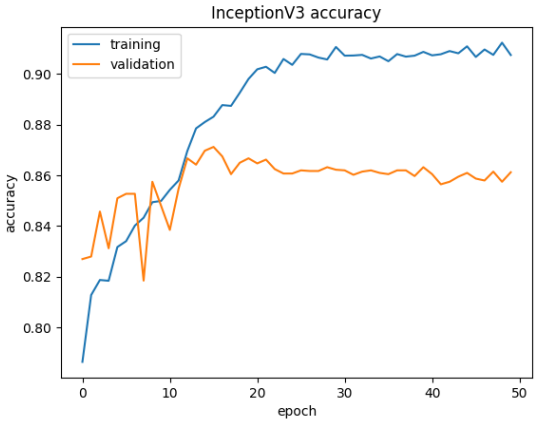

Figure 4: InceptionV3 training and validation accuracy.

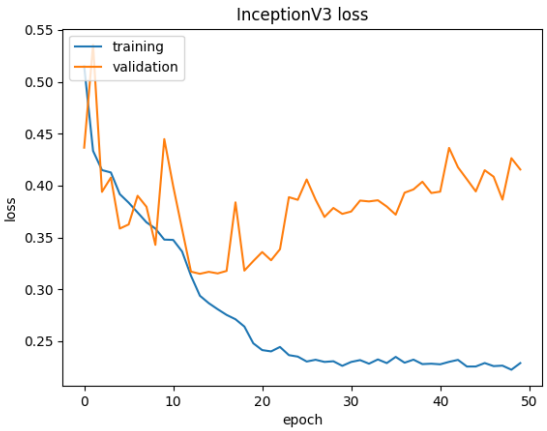

Figure 5: InceptionV3 training and validation loss.

3.2 *Discussion*

Based on Table 2, InceptionV3 achieved the best overall result performance. ResNet50V2 obtained a little lower performance with 74% training time. This verified that, although ResNet50V2 has more layers with more parameters, the shortcut connections were effective. Because of the disadvantage in result metrics was not significant compared to InceptionV3, it could be considered to use the small loss in accuracy in exchange for higher training efficiency to reduce time and cost. In this case, VGG16, which has the most parameters, does not have a good performance. It has the lowest accuracy and efficiency. In addition to this, it also possesses the most serious problem of overfitting.

The confusion matrix of Table 3 showed that although most of the samples were correctly predicted, a small percentage of the predictions were still problematic, especially false negative samples. For breast cancer, such accuracy is not sufficient to make CNNs the primary diagnostic method. However, the high efficiency of CNNs may allow them to become a supportive tool for breast cancer diagnosis. For example, to complete some preliminary examinations.

Overfitting was one reason that makes CNNs hard to increase the accuracy rate in this research. Since only 20,000 samples were extracted from the entire original dataset, the amount of data for training may not be sufficient. On the other hand, the three network architectures are based on higher resolution images, such as the ImageNet dataset with an average size of 469 × 387 pixels. While in this study, the original image sizes were all 50 × 50, so fewer effective features could be extracted from the images. This might also lead to accuracy problems.

4 CONCLUSIONS

This study proposed a neural network method by combining the cutting-edge CNN architectures with the diagnosing of invasive breast cancer. The main aim was to evaluate and compare the performance of ResNet50V2, VGG16, and InceptionV3 with metrices such as accuracy, recall, precision and F1-score. Experimental results showed that InceptionV3 and ResNet50V2 used in this research had a relatively good accuracy, while VGG16 had a lower performance with this case. The InceptionV3 had the best performance, with the accuracy of 87.12%. This network used the least training time with a similar accuracy. With the accuracy of these three networks, the CNN can be considered as a supportive tool at breast cancer diagnosing that may save pathologists much time. However, the dataset used in this research was a little limited. Small part of the original set was used, and image size limited the number of effective features in a single image. More architectures with better dataset should be evaluated in future works.

REFERENCES

[1] American Cancer Society (2022). Survival Rates for Breast Cancer. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

[2] American Cancer Society (2021). Invasive Breast Cancer (IDC/ILC). https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/invasive-breast-cancer.html

[3] Aslan, M. F., Sabanci, K., Durdu, A., & Unlersen, M. F. (2022). COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. Computers in Biology and Medicine, 105244.

[4] Bakator, M., & Radosav, D. (2018). Deep learning and medical diagnosis: A review of literature. Multimodal Technologies and Interaction, 2(3), 47.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.

[6] Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics, 7.

[7] Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., ... & Thoma, G. (2016). CNN-based image analysis for malaria diagnosis. In 2016 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 493-496). IEEE.

[8] Romano, A. M., & Hernandez, A. A. (2019, May). Enhanced deep learning approach for predicting invasive ductal carcinoma from histopathology images. In 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 142-148). IEEE.

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[11] World Health Organization (2021). World Cancer Day 2021: Spotlight on IARC research related to breast cancer. https://www.iarc.who.int/featured-news/world-cancer-day-2021/.