

# Convolutional Neural Networks for Breast Cancer Histology Image Classification Using Patch-Wise Training and Transfer Learning

Ian Timmis

Lawrence Technological University, Southfield MI 48075, USA  
itimmis@ltu.edu

**Abstract.** We developed an algorithm which can classify breast cancer histology images between four different categories achieving an accuracy exceeding highly specialized pathologists and any previous state of the art software. Those four categories being normal, benign abnormality, carcinoma in situ and invasive carcinoma. We ensemble an Inception-v3 Network, a 50-layer Residual Network and an Xception Network each pre-trained on ImageNet with custom logistic layers to define our network. We perform patch-wise training on the algorithm and then use a majority vote on the patch-wise predictions to output the image-wise prediction. We utilize modern deep learning concepts such as Transfer Learning to leverage the nuanced features learned from ImageNet. We achieve 97.5% accuracy on the test set.

**Keywords:** Deep Learning, Breast Cancer, Image Analysis, Convolutional Neural Networks, Transfer Learning, Artificial Intelligence.

## 1 Introduction

About 12.4% of U.S. women will develop invasive breast cancer in her lifetime. In 2018 it is estimated in the U.S. that 226,120 new cases of invasive breast cancer will be diagnosed in women and 2,550 new cases of invasive breast cancer will be diagnosed in men [1]. This does not include the estimated 63,960 cases of non-invasive (in situ) breast cancer that will be diagnosed in women this year. In the U.S. 40,920 women are expected to die in 2018 from breast cancer [1]. However, since 1989 there has been a decrease in the death rates. These decreases are thought to have partly occurred as the result of earlier detection [1]. As earlier detections have resulted in the decrease in death rates for almost three decades, it is clear that accurate classification is essential. Breast cancer can be diagnosed through the examination of biopsy tissue with hematoxylin and eosin (H&E) stained images. A study on U.S. pathologists showed an overall diagnostic concordance rate of 75.3% in which diagnostic interpretation was based on a single breast biopsy slide [2]. Interpreting histology images manually is a non-trivial task for specialized pathologists and motivate the use of Convolutional Neural Networks to discover patterns in the visual structure of these images for classification. In section 2, prior work done by computer vision and Artificial Intelligence experts is

introduced. Section 3 describes how our breast cancer image dataset is structured and preprocessed. Convolutional neural architecture of this study is explained in Section 4, followed by results and conclusion in section 5.

## 2 Prior work

Computer vision experts and machine learning practitioners have worked on automatically classifying H&E stained breast histology microscopy images for a long time. There are many examples of previous works attempting to do binary classification on this type of data. So far, these approaches have successfully reached over 90% accuracy on binary malignant-benign classification. Many of these have been outlined in the Araujo et al. paper [3]. In this paper, we are going to focus specifically on the Araujo et al. work as the prior work as we take this problem further and perform multiclass classification. Araujo et al. reports they achieved 77.8% accuracy for multiclass classification of the breast cancer histology images on two approaches they used to solve this problem. Each of their solutions utilized Convolutional Neural Networks (CNN) to solve this problem, however, one of them used a Support Vector Machine (SVM) as the final layer.

## 3 Input Dataset

The dataset described in this paper was released by the International Conference on Image Analysis and Recognition (ICIAR) 2018 challenge hosted by the Association for Image and Machine Intelligence (AIMI). The dataset provided includes 400 Hematoxylin and eosin (H&E) stained breast histology microscopy images. The large high definition images are of size  $1536 \times 2048$  pixels. They were acquired with magnification of 200x and pixel size of  $0.42\mu\text{m} \times 0.42\mu\text{m}$  [4]. The images are equally distributed among all classes. We use a 90/10 split between the train and test set. This results in a 360-image training set and a 40-image testing set.

### 3.1 Preprocessing

Before analyzing the data, the images need to be normalized. Data normalization is used to ensure that the input parameters have a similar data distribution. This leads to faster convergence during training. The images are normalized by applying the min-max normalization technique to the raw pixel values.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

The resulting pixel values are scaled to values between 0 and 1. The corresponding labels for each image are also transformed into categorical data through one-hot encoding.

### 3.2 Image Segmentation

The images are segmented into 48 separate image patches of size 256x256 (6 rows x 8 columns). There are two main motivations for this. The first motivation is to increase the dataset from 400 images to 19,200 images. The second motivation is for smaller image size training examples. Neural Networks have trouble classifying large high-resolution images. Conversely, Neural Networks are better equipped to recognize patterns in smaller pieces of data. Each of the image patches is labeled with the original image’s ground truth label. In theory, this may cause some patches to be mislabeled, but this will not harm performance and will be explained in further detail in section 4.4. The previous work segmented the images into 512x512 patches with 50% overlap. In our work, we discover that using patches of size 256x256 is adequate for finding relevant visual structures for high performance classification.

## 4 Method

### 4.1 Problem Formulation

The task of breast cancer detection is a multiclass classification problem, where the input is a 256x256 patch of a breast cancer histology image  $X$  and a label  $Y \in \{1, 2, 3, 4\}$  before one-hot encoding indicating whether the image contains normal tissue, benign abnormality, carcinoma in situ or invasive carcinoma respectively. For a single example in the training set, we minimize the categorical cross entropy loss

$$H(p, q) = - \sum_x p(x) \log(q(x)),$$

where  $p$  is the matrix containing the ground truth distributions and  $q$  is the matrix of approximate label distributions.

### 4.2 Network Architecture

The proposed algorithm utilizes 3 different CNN architectures. The 3 networks include an Inception-v3 Network [5], a 50-layer Residual Network [6] and an Xception Network [7]. Each of these networks are initialized with weights from networks pre-trained on the ImageNet dataset [8]. The three networks were specifically chosen because of their exceptional scores on the ImageNet dataset and the ease of setup using the Keras programming framework [9]. We remove the fully connected layer at the end of each of the network. The removed fully connected layer is then replaced with a 1024-neuron

fully connected layer using the Rectified Linear Unit (ReLU) activation function [10] followed by a 4 unit softmax layer for classification.

### 4.3 Training

To utilize the pre-trained weights learned from the ImageNet dataset, we utilize the idea of Transfer Learning [11]. We transfer useful knowledge from the ImageNet domain to the breast cancer classification domain through 2 steps of training.

In the first step of training, also called Transfer Learning, begins with “freezing” the original layers of the network so that the weights learned from ImageNet are not updated during backpropagation. The network is then trained using Adam optimization technique with standard parameters ( $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ ) [12].

The second step of training is called “Fine-Tuning.” During this stage we “unfreeze” several of the topmost layers and train the network using Stochastic Gradient Descent with Momentum using custom parameters ( $\alpha = 0.0001$ ,  $\gamma = 0.9$ ) [13]. The Inception-v3 Network, the 50-layer Residual Network and the Xception Network were unfrozen from the 249<sup>th</sup>, 140<sup>th</sup> and 126<sup>th</sup> layer respectively. It is important to note that these numbers are based off of the Keras implementation in which the total layer count includes layers that have non-learnable parameters (E.g. max pooling, batch normalization, dropout, etc.) as well as the layers with learnable parameters (E.g. convolutional and fully connected layers).

Each of the training steps utilized a mini-batch size of 16, model check pointing and early stopping. The Inception-v3 Network, the Xception Network and the 50-layer Residual Network achieved 84.9%, 84.84% and 85.1% accuracy respectively on patch-wise classification.

### 4.4 Ensemble

The 3 CNNs predictions are then input to our ensemble algorithm

$$p = (i + x + r) \odot \frac{1}{3},$$

where  $i$ ,  $x$  and  $r$  each represent a 4-dimensional vector containing the probabilities of each class prediction of the Inception-v3 Network, Xception Network and 50-layer Residual Network respectively.  $p$  is the resulting 4-dimensional patch-wise prediction vector. visual representation can be seen in Figure 1.

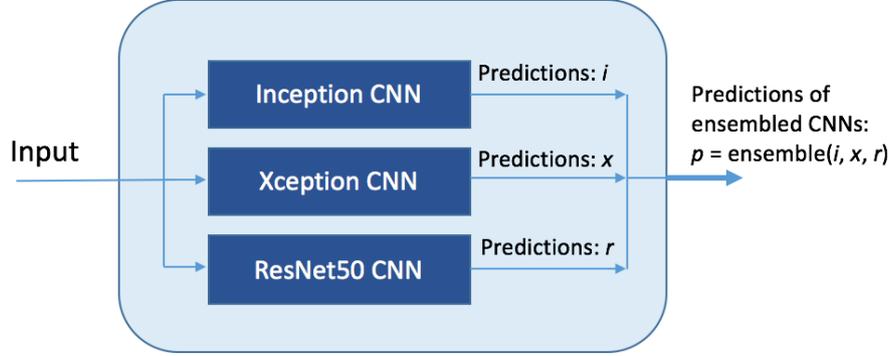


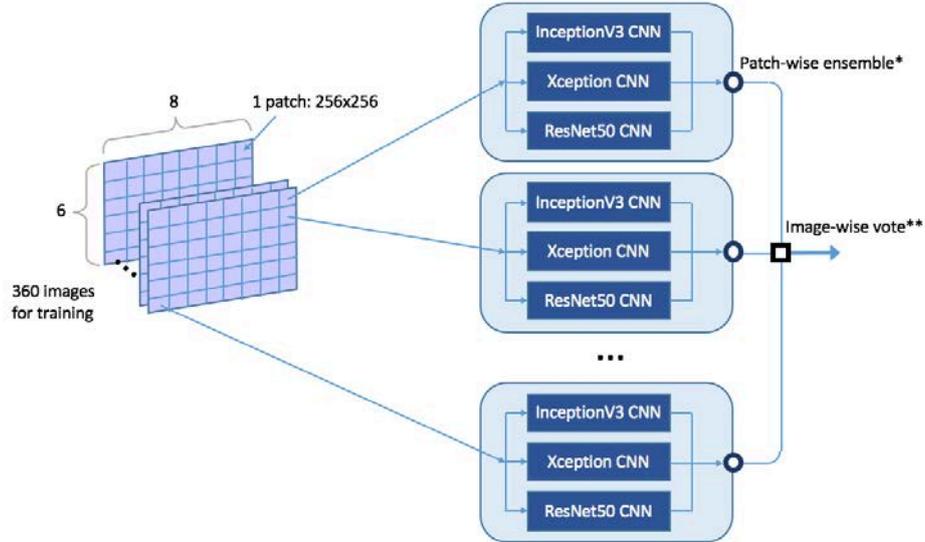
Fig. 1. The ensemble predictions calculated with the model predictions of each CNN

#### 4.5 Image-Wise Classification

Once all ensemble predictions for the 48 patches have been calculated, we then execute the majority vote algorithm on the ensemble predictions

$$\hat{Y} = \text{argmax}(\sum_{i=0}^{n_p} \text{onehot}(\text{argmax}(P_i))),$$

where P represents a  $4 \times 48$  dimensional matrix containing the patch-wise prediction vector for each patch of the image, and  $n_p$  represents the number of patches. The Image-wise classification process can be seen in Figure 2.



**Fig. 2.** The network comes to a consensus for full-image classification using a majority vote algorithm.

## 5 Results and Conclusion

The results of our algorithm are evaluated on classification accuracy. We determine the patch-wise classification accuracy for each of our individual networks as well as the Image-wise classification accuracy of our ensemble model. We compare our results directly with those of the prior work and can be seen in Table 1. Note that the patch-wise classification column represents 512x512 patches for the Arau'jo et al. experiments and 256x256 patches for our experiments.

We achieved an accuracy (97.5%) exceeding highly specialized pathologists and any previous state of the art software. We think this is due to (1) appropriate patch size, (2) ensemble of well-chosen CNN architectures and parameters, and (3) utilizing modern deep learning concepts such as Transfer Learning to leverage the nuanced features learned from ImageNet.

Future work for this project includes the analysis of any misclassification cases, testing on alternative consensus algorithms for image-wise classification other than the majority vote algorithm described earlier, testing with alternative image segmentation sizes and overlap, weighted ensembling based on patch-wise accuracy, the potential addition of other CNNs in the ensemble and further hyperparameter tuning.

**Table 1.** Accuracy of methods on the test set.

Research Group	Method	Patch-wise multiclass classification accuracy	Image-wise multiclass classification accuracy
Arau'jo et al.	CNN	66.7%	77.8%
	CNN with SVM	65%	77.8%
Timmis	Inception-v3	84.9%	<b>97.5%</b>
	Xception	84.84%	
	ResNet50	85.1%	

## 6 Acknowledgements

This project was started in Dr. Mazin Al-Hamando's class in the fall 2017 semester. The author would like to acknowledge Dr. CJ Chung his support reviewing the paper, designing the figures used throughout the paper, and covering the expenses for the cloud computing resources necessary for development to take place on a reasonable time scale.

## References

1. Breast Cancer Organization, [www.breastcancer.org/symptoms/understand\\_bc/statistics](http://www.breastcancer.org/symptoms/understand_bc/statistics), last accessed 2018/01/31.
2. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *Jama*. 2015; 313(11):1122. <https://doi.org/10.1001/jama.2015.1405> PMID: 25781441
3. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., ... & Campilho, A. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS one*, 12(6), e0177544.
4. ICIAR, <https://iciar2018-challenge.grand-challenge.org>, last accessed 2018/01/31.
5. 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).
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
7. Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. arXiv preprint [arXiv:1610.02357](https://arxiv.org/abs/1610.02357).
8. Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, and Fei-Fei, Li. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255. IEEE, 2009.
9. Chollet, F. Keras. <https://github.com/fchollet/keras>, 2015.
10. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).
11. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014, June). Learning and transferring mid-level image representations using convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on* (pp. 1717-1724). IEEE.
12. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
13. Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint [arXiv:1609.04747](https://arxiv.org/abs/1609.04747).