

Classifying Histopathology Images with Random Depthwise Convolutional Neural Networks

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

Abstract—

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

This document is a model and instructions for L^AT_EX. Please observe the conference page limits.

II. METHODS

we will look at classifying the breast cancer biopsy image set first as two classes (cancer [class 1] or no cancer [class 0]), then later consider the specific anomalies in the dataset and label the images in a multi-class configuration. we will label the images in the following seven classes:

a. Ductal Carcinoma [class 0, malignant], b. Lobular Carcinoma [class 1, malignant], c. Mucinous Carcinoma [class 2, malignant], d. Papillary Carcinoma [class 3, malignant], e. Fibroadenoma [class 4, benign], f. Phyllodes Tumor [class 5, benign], g. Tubular Adenoma [class 6, benign].

Figure 1 shows a sample of the images in each of the seven classes we defined. In the two class machine learning experiments, images are classified as cancerous or non-cancerous.

A. Breast Cancer Tissue Biopsy Image Pre-Processing and Preparation

The breast biopsy images will be kept in their original 3-channel PNG format, and since they are already uniform in size, the images are simply downsized and stored in NumPy arrays. Each image is flattened out and converted to a feature vector. At 350x230 pixels, the feature vector will have 80,500 features, reducing memory requirements and training times. The vectors are then normalized if needed or required by the classification method.

In order to train and evaluate our classifier, a label vector that stores the true class of each data sample (e.g., [0,1] for two classes, [0,1,2,3,4,5,6] for seven classes) is created and used as a true label input along with the sample data. We

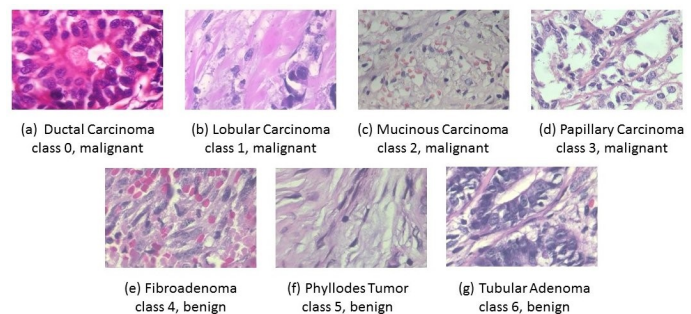


Fig. 1. Sample biopsy images showing the 7 anomalies being classified. (a) Ductal Carcinoma [class 0, malignant], (b) Lobular Carcinoma [class 1, malignant], (c) Mucinous Carcinoma [class 2, malignant], (d) Papillary Carcinoma [class 3, malignant], (e) Fibroadenoma [class 4, benign], (f) Phyllodes Tumor [class 5, benign], (g) Tubular Adenoma [class 6, benign].

note that for the non-augmented set, 1,232 samples are labeled as cancerous, and 374 samples as non-cancerous for a total of 1,606 samples. Given that the data is imbalanced in favor of one of the cancer types, we augment the data to create a more balanced set. Augmentation is done only 1-fold on the most common class (ductal carcinoma) and 6-fold on all other classes. A combination of horizontal and vertical flip, rotation (180 degrees of rotation is used to maintain the shape of the image) and shifting is performed on the resized images to create more samples.

B. VGG16 & RDCNN

Convolutional Neural Networks are typically composed of alternating convolution and pooling layers followed by a final flattened layer. A convolution layer is specified by a filter size and the number of filters in the layer. Each convolution layer performs a moving dot product against pixels given by a fixed filter of a predetermined size. The dot product is made non-linear by passing the output to an activation function

such as a sigmoid or ReLU (Rectified Linear Unit) function. Two convolutional neural networks are used to classify the breast cancer images. The popular VGG-16 network and the recently developed Random Depth-wise Convolutional Neural Network (RDCNN). This new network attempts to learn a feature space with random depth-wise convolutions on which a linear support vector machine or stochastic gradient descent is then applied. The figures below illustrate both networks and give a brief description of each.

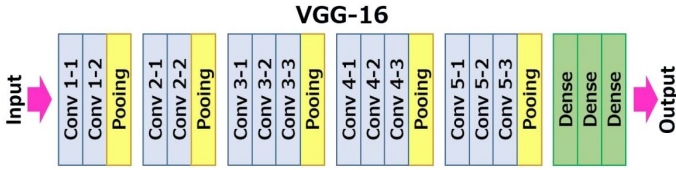


Fig. 2. Flow diagram of the VGG-16 convolutional neural network. It contains 13 convolutional layers (with ReLU activation), 5 pooling layers and 3 fully connected layers (with ReLU and Softmax activation). The model was proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition.” Source: *Muneeb ul Hassan*. <https://neurohive.io/en/popular-networks/vgg16/> (accessed in August 2019)

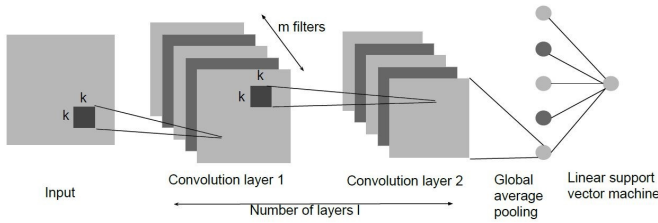


Fig. 3. A random depth-wise convolutional neural network with two layers. Diagram shows filter size of k , and $m = 5$ filters in each layer. Source: *Yunzhe Xue and Usman Roshan*. <http://scinapse.io/papers/2808187103> (accessed in August 2019)

Our CNN experiments are conducted in Python using the Keras library implementations for VGG-16. The loss function used is Categorical Cross-entropy (since it generated the highest accuracy after many trials), and the SGD (Stochastic Gradient Descent) optimizer is selected, as it generated the best results with a learning rate of 10^{-4} . The networks are trained using an increasing number of epochs, starting with 10 epochs, up to 50 epochs, in steps of 10. For VGG-16, the following parameters were used: number of filters = 64, 128, 256, 512; kernel dimension = 3×3 pixels; max pooling = 2×2 pixels. For RDCNN, a TensorFlow implementation is used (faster than Keras). We experimented with the following parameters: models: STL10, Cifar10; number of features: 2500, 10000; iterations: 5000; structures: 7 layers, 25 layers. In addition to using the full x-ray images, we also used cropped tissue images by removing most of the irrelevant background. This reduces the image size and makes the images more uniform.

C. Description of Breast Cancer Biopsy Data

The microscopic biopsy images in our BreakHis dataset were collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). The images are provided in their raw PNG (Portable Network Graphic) format, without normalization or color standardization and are all the same size (700x460 pixels, 3-channel RGB, 8-bit depth per channel). The dataset is divided into two main groups: benign tumors and malignant tumors. A lesion is referred to as histologically benign when it does not match any criteria of malignancy. Malignant tumors are cancerous lesions that can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastasize) to cause death. The samples present in this dataset were collected by SOB (Surgical Open Biopsy) method, also called partial mastectomy or excisional biopsy. This type of procedure removes a large tissue sample and is done in a hospital with general anesthesia.

The benign and malignant groups are further divided into sub-groups describing the specific kind of anomaly. For benign lesions, the anomalies present are fibroadenoma, Phyllodes tumor and tubular adenoma. For the malignant lesions, the anomalies present are ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma. In our experiments, we will only consider images at the 400X magnification level, where we count a total of 1,606 samples. Out of that total, 374 samples are benign and 1,232 are malignant. We can see that this is a greatly imbalanced dataset in favor of malignant tumors.

D. Breast Cancer Tissue Biopsy Dataset Exploration

The BreakHis images are relatively large in their native format. Each image is about about 500 KB on disk and has a uniform size of 700x460 pixels, which would generate a feature vector of 315,000 features. To make these images more manageable, we downsized them to 350x230 pixels. This will require a lot less memory and less compute resources. Additionally, we note that the dataset contains about 6 times more ductal carcinoma images than all other classes (average of 136 samples per class). We will use augmentation (adding more samples by rotating and flipping the original images) to balance the samples, and we will test with both non-augmented and augmented datasets to evaluate and compare classifier performance.

E. Breast Cancer Tissue Biopsy Image Pre-Processing and Preparation

The breast biopsy images will be kept in their original 3-channel PNG format, and since they are already uniform in size, the images are simply downsized and stored in NumPy arrays. Each image is flattened out and converted to a feature vector. At 350x230 pixels, the feature vector will have 80,500 features, reducing memory requirements and training times. The vectors are then normalized if needed or required by the classification method.

In order to train and evaluate our classifier, a label vector that stores the true class of each data sample (e.g., [0,1] for

two classes, [0,1,2,3,4,5,6] for seven classes) is created and used as a true label input along with the sample data. We note that for the non-augmented set, 1,232 samples are labeled as cancerous, and 374 samples as non-cancerous for a total of 1,606 samples. Given that the data is imbalanced in favor of one of the cancer types, we augment the data to create a more balanced set. Augmentation is done only 1-fold on the most common class (ductal carcinoma) and 6-fold on all other classes. A combination of horizontal and vertical flip, rotation (180 degrees of rotation is used to maintain the shape of the image) and shifting is performed on the resized images to create more samples. Figure 2 shows the augmented versions of a biopsy sample.

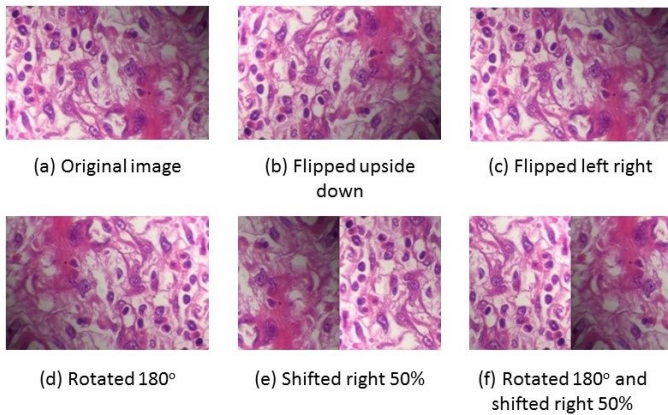


Fig. 4. Sample biopsy image with various augmentation transformations. (a) Original resized image, (b) Flipped upside down, (c) Flipped left/right, (d) Rotated 180°, (e) Shifted right 50%, (f) Rotated 180° and Shifted right 50%.

III. RESULTS

We present the breast cancer biopsy image classification results. For both 2-class and 7-class datasets, the RDCNN neural network did best, followed by VGG-16, then Random Forest, Logistic Regression and SVM. Detailed results for the most significant experiments are given in the sections below.

A. Binary class

RDCNN performed very well in the two-class experiments, reaching a validation accuracy of over 92%. VGG did relatively well at about 80% accuracy. Random Forest, SVM and Logistic Regression came in last at around 75%. The biopsy images are clearly a lot easier to classify than the breast x-rays. Surprisingly, SVM and Logistic Regression did worse on the augmented data, but all other classifiers did better with 8,120 total samples. Training accuracies were very high, but the validation accuracies fell short for SMV and Logistic Regression. RDCNN did even better with the augmented dataset, with the accuracy reaching 98%. Although training times for RDCNN were in the order of minutes, it took about 20 hours to generate the new features and then train the Linear SVC classifier. The table below outlines the most notable results from the many experiments that were run on this dataset.

TABLE I
TWO-CLASS BREAST BIOPSY IMAGE CLASSIFICATION RESULTS – NO AUGMENTATION

Input Data	Classifier	Training Accuracy	Validation Accuracy	Training Time [hours]
230x350 PNG	SVM	100%	75.64%	0.08
230x350 PNG	Logistic Regression	100%	75.43%	0.13
230x350 PNG	Random Forest	100%	74.44%	0.17
230x350 PNG	VGG16	89.41%	80.49%	1
230x350 PNG	RDCNN	100%	92.55%	0.04

TABLE II
SEVEN-CLASS BIOPSY IMAGE CLASSIFICATION RESULTS – NO AUGMENTATION

Input Data	Classifier	Training Accuracy	Validation Accuracy	Training Time [hours]
230x350 PNG	SVM	98.60%	43.79%	0.13
230x350 PNG	Logistic Regression	98.60%	43.79%	0.25
230x350 PNG	Random Forest	98.60%	53.42%	0.13
230x350 PNG	VGG16	98.60%	52.48%	0.5
230x350 PNG	RDCNN	98.60%	67.08%	0.28

B. Multi class

In the seven-class experiments, it is clear that the classes are harder to detect. The accuracies obtained were below 50%, with the exception of Random Forest which yielded a validation accuracy of about 53%, and RDCNN which reached a 67% accuracy on the original dataset (1,606 samples). SVM and Logistic Regression did worse on the augmented dataset, but the rest of the classifiers did considerably better. Random Forest went from 53% to 63%. VGG-16 went from about 46% to about 71%. RDCNN accuracy increased from 67% to about 95%. New feature generation took about 25 hours. Again, we note the very long training times for the deep learning methods (VGG16 and RDCNN if we include the feature generation time). The tables below display the most notable results obtained.

C. Effect of data augmentation

IV. DISCUSSION

ACKNOWLEDGMENT

REFERENCES

TABLE III
TWO-CLASS BREAST BIOPSY IMAGE CLASSIFICATION RESULTS –
AUGMENTED DATA

Input Data	Classifier	Training Accuracy	Validation Accuracy	Training Time [hours]
230x350 PNG	SVM	100%	70.83%	1.5
230x350 PNG	Logistic Regression	100%	71.53%	0.3
230x350 PNG	Random Forest	100%	78.76%	2
230x350 PNG	VGG16	82.81%	81.78%	0.5
230x350 PNG	RDCNN	100%	98.77%	0.12

TABLE IV
SEVEN-CLASS BIOPSY IMAGE CLASSIFICATION RESULTS – AUGMENTED
DATA

Input Data	Classifier	Training Accuracy	Validation Accuracy	Training Time [hours]
230x350 PNG	SVM	99.49%	35.45%	4
230x350 PNG	Logistic Regression	99.49%	35.63%	3
230x350 PNG	Random Forest	99.49%	63.05%	0.5
230x350 PNG	VGG16	99.14%	70.61%	7.5
230x350 PNG	RDCNN	99.49%	95.38%	1