
Final Project: Breast Cancer Detection in Histopathological Images using various Convolutional Neural Networks

Jasmine Batra

Department of Computer Engineering
New York University
New York, NY 11201
jb7854@nyu.edu

Sachin Karthik Kumar

Department of Computer Engineering
New York University
New York, NY 11201
sk9622@nyu.edu

Rudrasinh Ravalji

Department of Computer Engineering
New York University
New York, NY 11201
rnr8237@nyu.edu

Abstract

Cancer is one of the world's most frequent and deadly diseases. One in every eight women and one in every eight hundred males is diagnosed with breast cancer. As a result, our first goal should be early cancer detection, as early detection can aid in the effective treatment of cancer. As a result, we present a saliency detection method based on sophisticated deep learning techniques, in which the machine is taught to mimic pathologists' movements for the localisation of diagnostically relevant regions. We use a CNN to train to identify diagnostic types of breast cancer (VGG16, EfficientNet, ResNet architecture). To train our model, we used the BreakHis dataset. In histopathology imaging, we focus on both detecting and classifying malignant areas. The 'salient' areas are those that are diagnostically important. Pathologists and medical institutions will be able to use the detection technology as an open-source online application.

[Github](#)

1 Introduction

Cancer is one of the most frequent diseases that a person can contract, and it is also the second leading cause of mortality. Cancer cells divide and spread in the surrounding tissues in any part of the body and in all of its forms. Breast cancer is one of the most serious and prevalent cancers in women, and it can be caused by a variety of factors including lifestyle, screening, and family history [1]. Breast cancer affects one in every eight women and one in every 800 males, according to statistics. Breast exams, mammograms, ultrasounds, collecting a sample of breast tissue (biopsy), and breast magnetic resonance imaging are just a few of the procedures to test and diagnose breast cancer. When compared to high-income countries, the number of cases of breast cancer is increasing by the day in low and middle-income countries. One of the major reasons for this is the high cost of getting yourself diagnosed, as well as the substandard treatment provided in low and middle-income countries. Many of the resources needed for mammography diagnosis are not available in low- and middle-income countries [2]. Due to a lack of resources, the death rate from breast cancer has not decreased with diagnosis from clinical breast examination or self-examination. Even though we live in a time where digitization is prevalent in the field of diagnosis, pathology remains largely reliant on investigations conducted using microscopic examinations of tissues on glass slides. The advancement of machine vision and digitally scanning microscopes has enhanced the possibility of computer-assisted diagnosis in recent years, allowing slides of tissue histology to be recorded as

digital images. As a result, image analysis and machine learning approaches have benefited from digital tissue histology.

2 Background Study and Related Works

2.1 Literature Survey

There hasn't been much research done in this area previously. Though there have been a lot of new approaches recently. This is a new field with a lot of room for advancement. As a result, our effort intends to make a substantial contribution to the discipline.

Breast Cancer Diagnosis Using Deep Learning Algorithm [1] is one of the few publications that uses the Wisconsin Breast Cancer Database. The database in question has 569 rows and 30 characteristics. For scaled datasets, they used pre-processing methods such as normalizers and label encoders. In 2018, their work was published. Breast Cancer Diagnosis from Histopathological Image based on Deep Learning is another work that leverages the BreakHis dataset. CNNs, or convolutional neural networks, were utilized to identify images as benign or cancerous. However, they employed the Inception V3 model, which has a poor initialization rate. A lot of computation time is wasted and adjusting is costly.

In the year 2019, such material was published. Nuclear Atypia Grading in Histopathological Images of Breast Cancer [2] is another paper. CNN-based approach was utilized with Convolutional Neural Networks. They graded mitotic cells, nuclear atypia, and tubule development using Nottingham's technique, which has three elements. Their model is made up of two parts: a feature extraction section with convolutional layers and an activation function section with ReLu and pooling layers.

In 2018, the same was published. Another approach that leverages the BreakHis dataset is Histopathological Image Analysis for Breast Cancer Detection Using Cubic SVM [3]. They examined six SVM variations, including support vector machine, random forest technique, and KNN (K Nearest Neighbors), for experimental purposes. They discovered that cubic SVM outperformed all of the other approaches. When the class size is big, however, SVM struggles to predict class labels. In the year 2020, this work was released.

Another technique that employed BreakHis dataset was A Deep CNN Technique for Detection of Breast Cancer Using Histopathology Images[4]. They also employ CNN, or convolutional neural networks, to differentiate between cancerous and benign images. Breast cancer is the most prevalent cancer diagnosed in women in the United States (excluding skin cancers), accounting for 30% of all new cancer diagnoses in women. Histopathology tries to differentiate between normal tissue, benign (benign) and malignant (carcinomas) abnormalities, as well as perform a prognosis evaluation.

2.2 Dataset

There are a variety of histopathological imaging data sets that may be used with deep learning approaches to detect tumor tissue. We used the BreakHis dataset, which consists of 9109 microscopic images of breast tumor tissue from 82 individuals using various magnification factors (40X,100X,200X,400X). It is divided into two types: benign and malignant. Tumor tissue is missing in benign tumors, but tumor tissue is present in malignant tumors. Figure 1 shows the structure of the BreakHis Dataset. There are now 2,480 benign and 5,429 malignant samples in the database (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format). This database was created in partnership with Parana, Brazil's P&D Laboratory - Pathological Anatomy and Cytopathology.

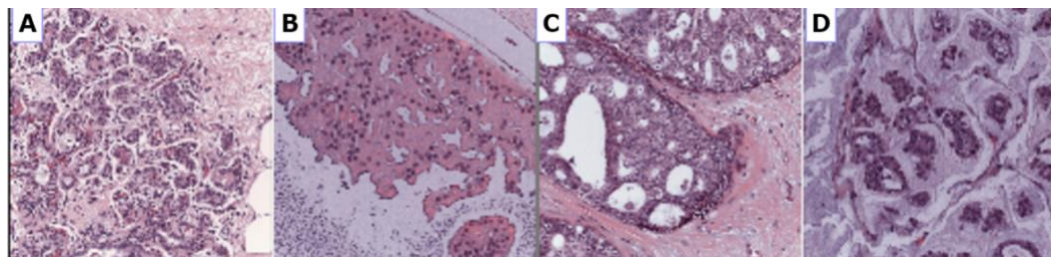


Figure 1: Example of images available in BreakHis Dataset.

3 Methodology

3.1 Current Work

The Initial Stage of our project involved by uploading the dataset to google drive and running it on google collab. We couldn't access the dataset from Grand Challenge website. We asked them for the access of the dataset, and they sent us the download link. After researching a bit, we went with the BreakHis dataset, which was gathered in 2014 by P&D Lab in Brazil. The dataset that was provided to us had two folders benign and malignant images. So, we split our data into train and valid. So, our dataset is prepared as dictionaries rather than list. This way we link images in our training process. With google collab, it was taking too much of processing power to prepare our training set given the number of images in our dataset. This was the reason that we made our dataset available on our HPC Greene account. Thus, we were able to upload the data. We applied data augmentation on the data After that data got ready, so the next step to build and train the classifier so we decided used pretrained ResNet 152 to extract features for classification. The main merit of using ResNet 152 is that it produces better classification accuracies without increasing the complexity of the model. he proposed deep learning technique ResNet-152 is a useful and responsible approach in comparison to traditional methods In this work, ResNet-152 deep learning structure is used for feature extraction and classification. Compared to other network models, deeper ResNet have lower training errors, optimization, and generalization ability. It allows for very smooth. forward and backward propagation, making it much easier to optimize deeper models, and deeper address generalization. We tried training our model on different optimizers and different learning rates. We even tried to work on predicting the class of the cancer but due to some code errors we couldn't execute that part.

Table 1: Accuracies on different optimizers and different learning rate

Optimizer	Learning rate	Accuracy
SGD	0.1	83.91
SGD	0.01	86.45
SGD	0.001	87.33
RMSprop	0.1	81.68
RMSprop	0.01	85.79
RMSprop	0.001	84.53
Adam	0.1	72.48
Adam	0.01	76.23
Adam	0.001	79.52

- Dropout: 0.5
- Epochs: 100
- Optimizer: Adam
- Batch size: 32
- Learning rate: 0.1,0.01,0.001
- First dense layer has 2048 dimensions
- Removed Dropouts in the convolution layers
- Reduce pool size for Better and finer patterns

3.2 Next Steps

Now that our model trained, we will be predicting which class of cancer it is and the model can be improved by changing some of the hyperparameters. We will even using different pretrained model such as Efficient Net, VGG16 and compare at the end to know which model improves the accuracy. Our future work involves deploying a website where a user can put his/her details and get a result based on image input.

References

- [1] N. Khuriwal and N. Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 98-103.
- [2] S. K. Jafarbiglo, H. Danyali and M. S. Helfroush, "Nuclear Atypia Grading in Histopathological Images of Breast Cancer Using Convolutional Neural Networks," 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), 2018, pp. 89-93.

135 [3] S. Singh and R. Kumar, "Histopathological Image Analysis for Breast Cancer Detection Using
136 Cubic SVM," 2020 7th International Conference on Signal Processin
137 [4] G. Wadhwa and A. Kaur, "A Deep CNN Technique for Detection of Breast Cancer Using
138 Histopathology Images," 2020 Advanced Computing and Communication Technologies for High
139 Performance Applications (ACCTHPA), 2020, pp. 179-185[7]
140 [5] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer
141 Histopathological Image Classification," IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–
142 1462, Jul. 2016.
143