BREAST CANCER DETECTION

An Image Classification CNN Pytorch Model

Team Members

Ahmed Hamed Akef

Abdulrahman Issam AbouOuf

Khaled Abdulkader Madkour

Mohamad Nasser ElFarash

Supervisor

Dr. Amira Yassin,PhD

TABLE OF CONTENTS

TABLE OF FIGURES

TABLES

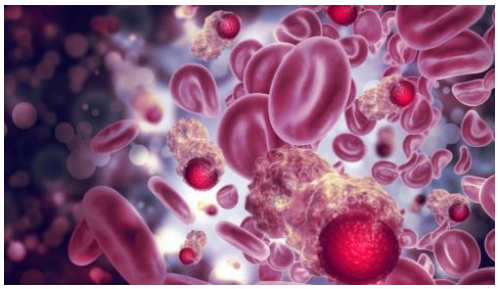
1 INTRODUCTION

This paper will introduce the project of Neural Network and how did we use it to solve the problem of Breast Cancer detection.

It will discuss how did we train the network, what architectures did we use and what is the dataset we used.

1.1 CANCER AND BREAST CANCER

Cancer is a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body.



Breast cancer is a type of cancer that starts in the breast. Cancer starts when cells begin to grow out of control.

Breast cancer cells usually form a tumor that can be seen on an **x-ray** or felt as a lump. Breast cancer occurs almost entirely in women, but men can get breast cancer, too.

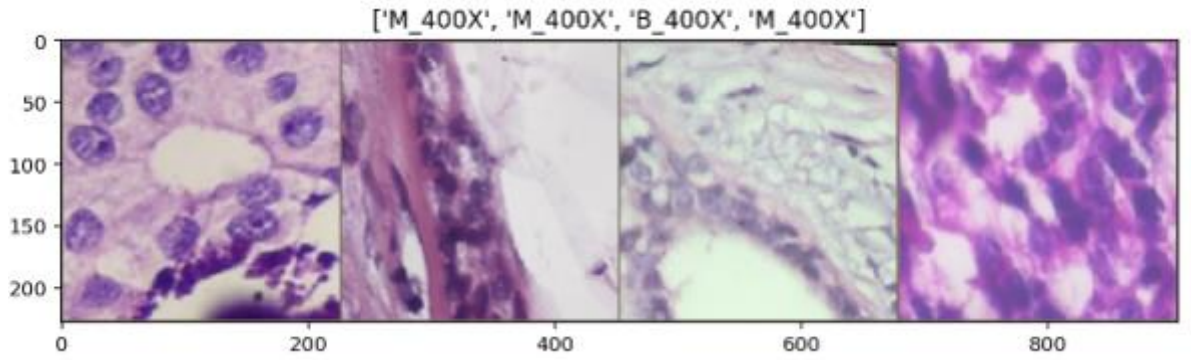


Figure 1-1 Breast Cancer sample from our Dataset

2 PROBLEM STATEMENT

Breast cancer can spread when the cancer cells get into the blood or lymph system and are carried to other parts of the body.

If cancer cells have spread to your lymph nodes, there is a higher chance that the cells could have traveled through the lymph system and spread (metastasized) to other parts of your body.

2.1 STATISTICS

Breast cancer is the most commonly occurring cancer in women and the second most common cancer overall. There were over 2 million new cases in 2018. The top 25 countries with the highest rates of breast cancer in 2018 are given in the table below.

34% of Egyptian women suffer from breast cancer, the most common type of cancer in Egypt for women.

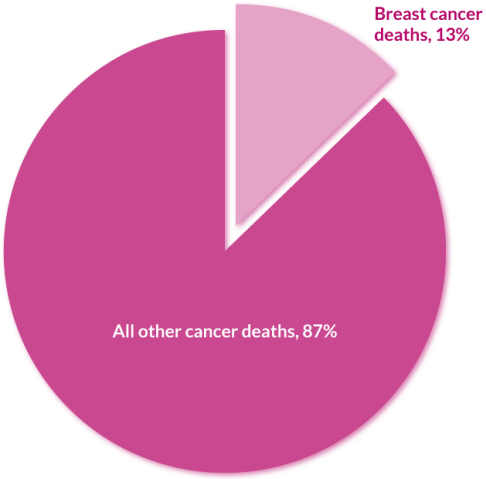


Figure 2-1 Estimated Cancer Deaths in Women in 2019

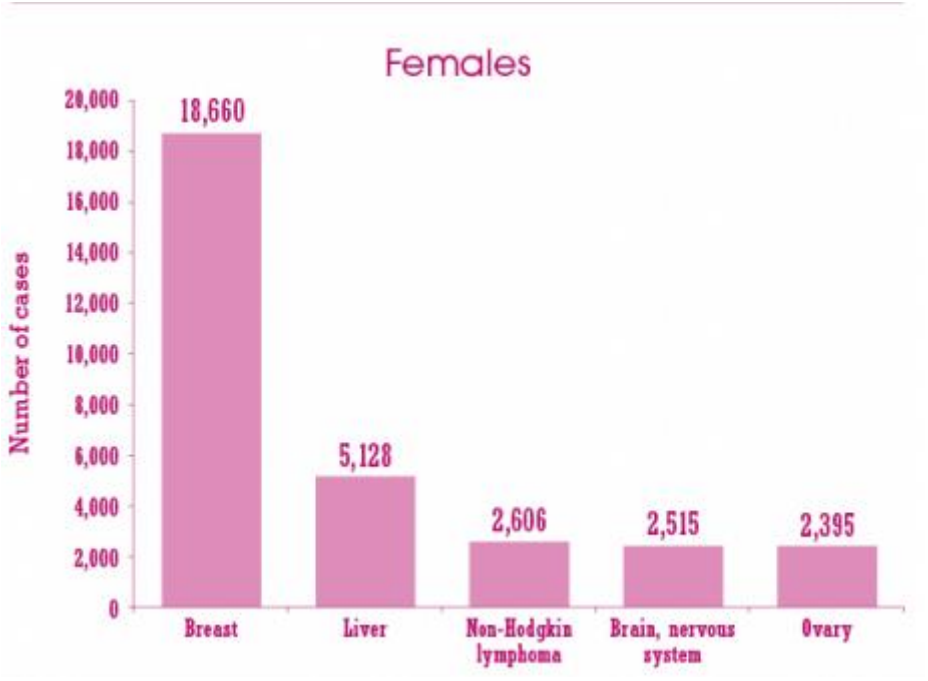


Figure 2-2 Female Cancer Statistics in Egypt

3 OUR SOLUTION

Our project introduces Neural Networks (Deep Learning) to try to detect if the patient is Benign or Malignant. We used ResNet-50 as the best architecture after testing (more on that later). And we used [Breast Cancer Histopathological Database (BreakHis)] dataset.

3.1 DATASET

The Breast Cancer Histopathological Image Classification (BreakHis) is composed of 9,109 microscopic images of breast tumor tissue collected from 82 patients. it contains 2,480 benign and 5,429 malignant samples (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format).

The dataset BreaKHis is divided into two main groups: benign tumors and malignant tumors.

Both breast tumors benign and malignant can be sorted into different types based on the way the tumoral cells look under the microscope.

Magnification	Benign	Malignant	Total
40X	652	1,370	1,995
100X	644	1,437	2,081
200X	623	1,390	2,013
400X	588	1,232	1,820
Total of Images	2,480	5,429	7,909

Table 1 Dataset

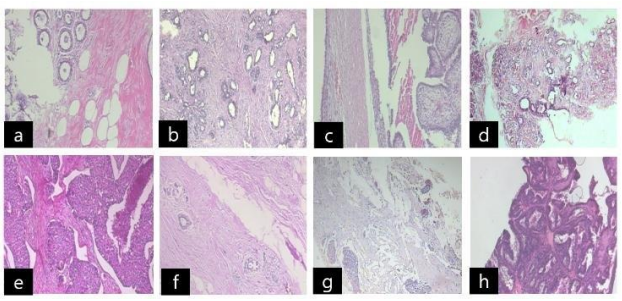


Figure 3-1 Dataset Sample

3.2 ARCHITECTURE

ResNet-50 freezing 20 layer was the best accuracy we experienced.

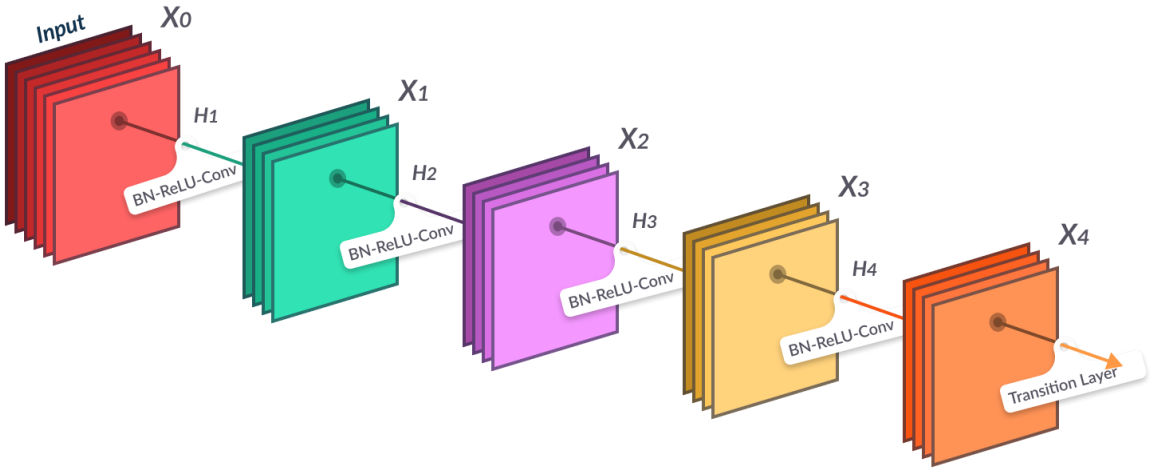


Figure 3-2 ResNet Block

3.3 DATA AUGMENTATION

Choosing the right defect on your images as augmentation[1] can improve your performance we did some experiences to figure out what would lead us to better model evaluation.

3.3.1 DATA AUGMENTATION IN MEDICAL DATASETS

Any medical dataset augmentation (defection) is very critical issue. As you can't defect the image in any way that can mislead info. For example, colorjitter can mislead the skin or the lump to get real info.

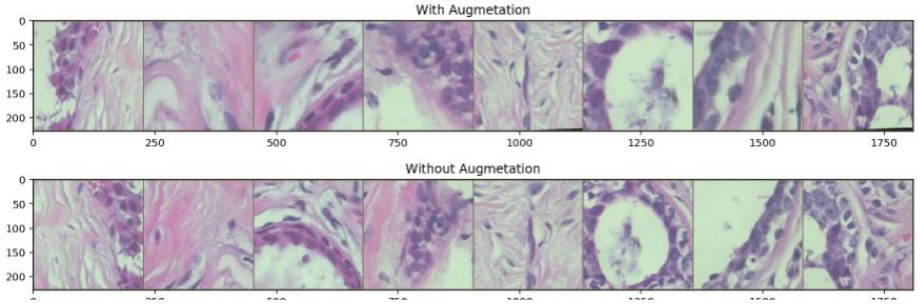


Figure 3-3 Data with and without Augmentation

[1] Data augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data like resizing, color change, random image flip, etc.

```
data_transforms = {
    'train': transforms.Compose([
        transforms.RandomRotation(5),
        transforms.Resize(224),
        transforms.RandomResizedCrop(224),
        #transforms.ColorJitter(brightness=0.4, contrast=0.4, saturation=0.4),
        transforms.RandomHorizontalFlip(),
        transforms.ToTensor(),
        transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
    ]),
    'test': transforms.Compose([
        transforms.Resize(224),
        transforms.CenterCrop(224),
        transforms.ToTensor(),
        transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
```

Figure 3-4 Augmentation Code

resizing the images to be faster to process and flipping to increase data.

3.4 BUILDING AND TRAINING THE CLASSIFIER

	Architecture	Accuracy	Notes
Experiment 1	ResNet 152-Layers	78%	Layers all layer freezed
Experiment 2	ResNet 18-Layers	81%	No layer freezed
Experiment 3	ResNet 50-Layers	82%	No freezing
Experiment 4	ResNet 50-Layers	76%	25 layers freezed no resizing – Very Slow
Experiment 5	ResNet 50-Layers	88%	20 layers freezed with resizing

Table 2 Architecture Experiments

3.5 MODEL EVALUATION

Recall	Precision	F1 Score[2]
0.8086	0.8378	0.8193

Table 3 Model Evaluation

Class	M	B
M	165	72
B	33	385

Table 4 Confusion Matrix

The confusion matrix[3] shows that it matches the most pictures. M for Malignant and B for Benign.

4 MODEL RESULTS

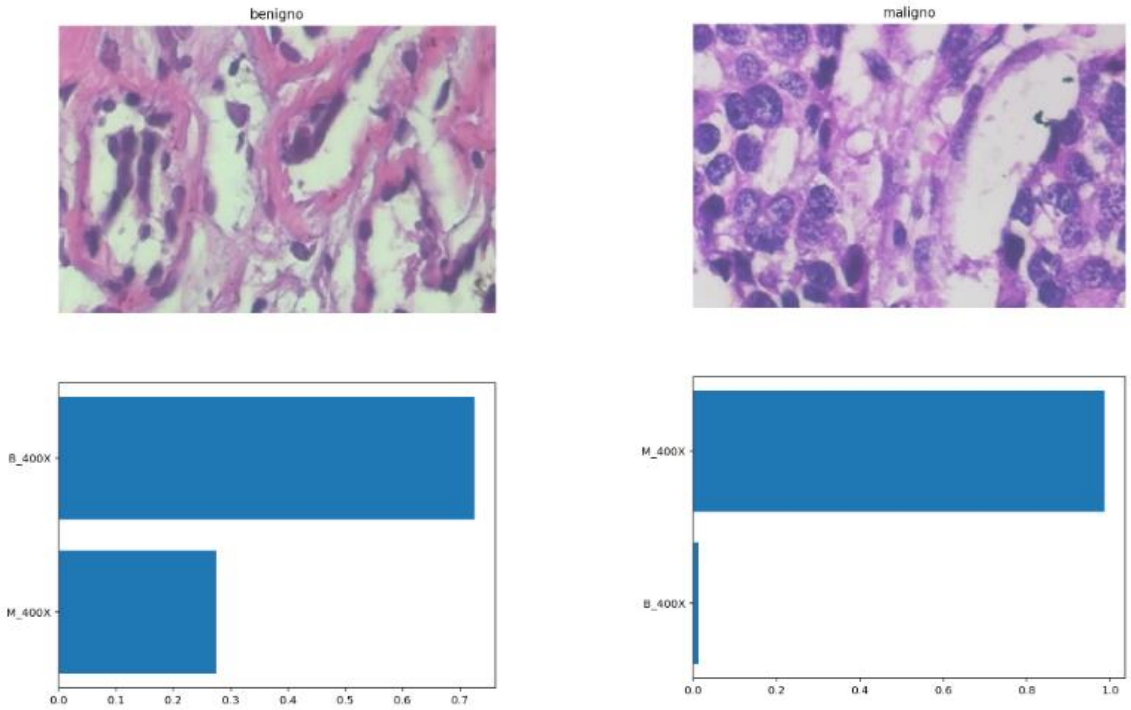
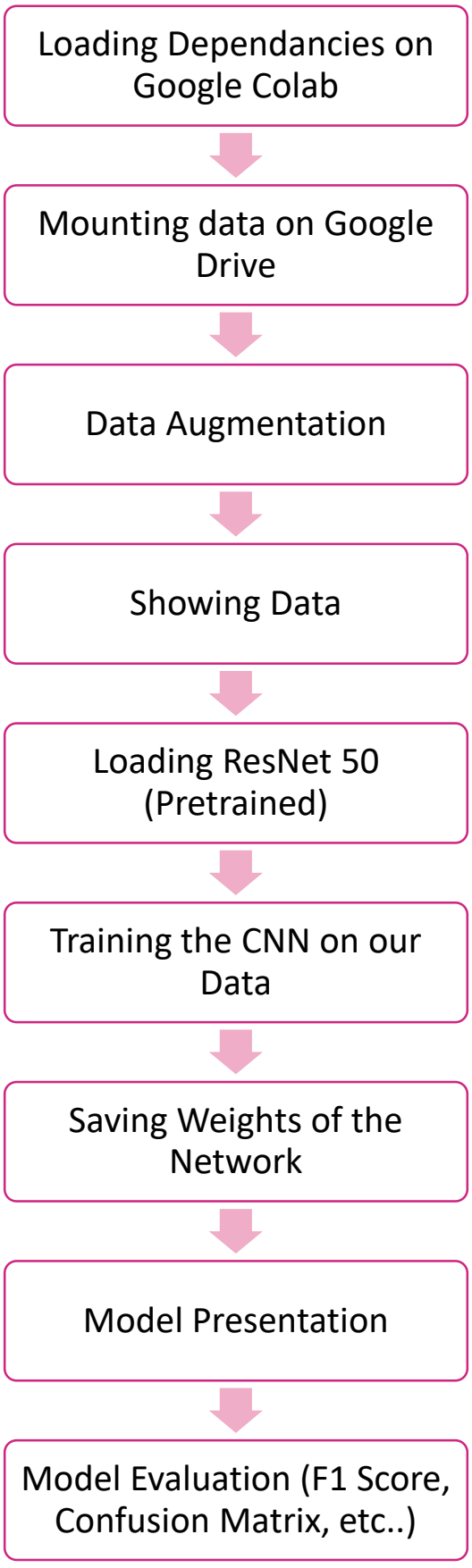


Figure 4-1 Results Of our Model

[2] the F_1 score is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.

[3] A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

5 MODEL FLOWCHART



6 REFERENCES

- Breast Cancer Histopathological Database (BreakHis)
- Breast Cancer Facts & Figures 2019-2020
- Cancer In Egypt
- https://www.cancer.net/cancer-types/breast-cancer/statistics
- https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics
- https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/