Breast Cancer detection

Rotem Feinblat, Hadar Pur

Submitted as final project report for the Deep Learning course,
IDC, 2020

1 Introduction

Cancer is a fatal illness often caused by genetic disorder aggregation and a variety of pathological changes. Cancerous cells are abnormal areas often growing in any part of human body that are life-threatening. Cancer also known as tumor must be quickly and correctly detected in the initial stage to identify what might be beneficial for its cure. Even though modality has different considerations, such as complicated history, improper diagnostics and treatment that are main causes of deaths. The first goal of the project is to detect breast cancer using deep learning technique. Second goal of the project was to train a Neural network on breast cancer and check how it affect on the performance of other type of cancer. Finally, challenges are also highlighted for possible future work.

1.1 Related Works

There are many exiting works that can be used as a reference for cancer detection. [1] [2]

2 Solution

2.1 General approach

We will try to detect Breast cancer from data set by implementing CNN. First,we will try to improve the performance of the net by changing the number of layers and the parameters itself. Second, we would like to check if a trained net can perform better results for other type of cancer.

2.2   Data set

the following images represent each class from the data set

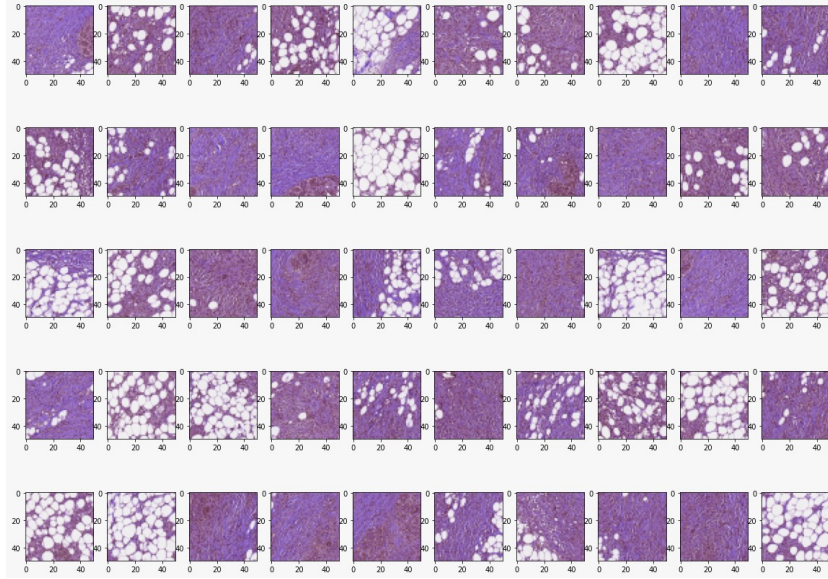Figure 1: Malignant Breast cancer
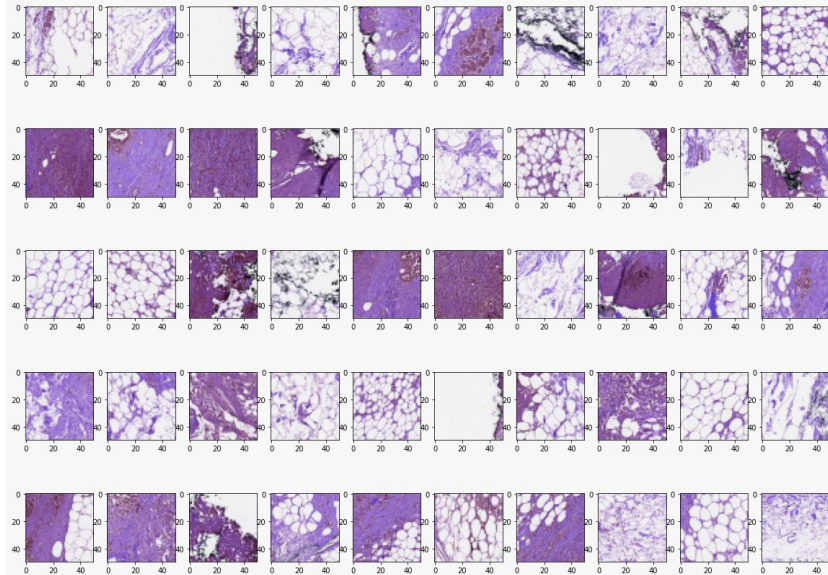

Figure 2: Benign Breast cancer
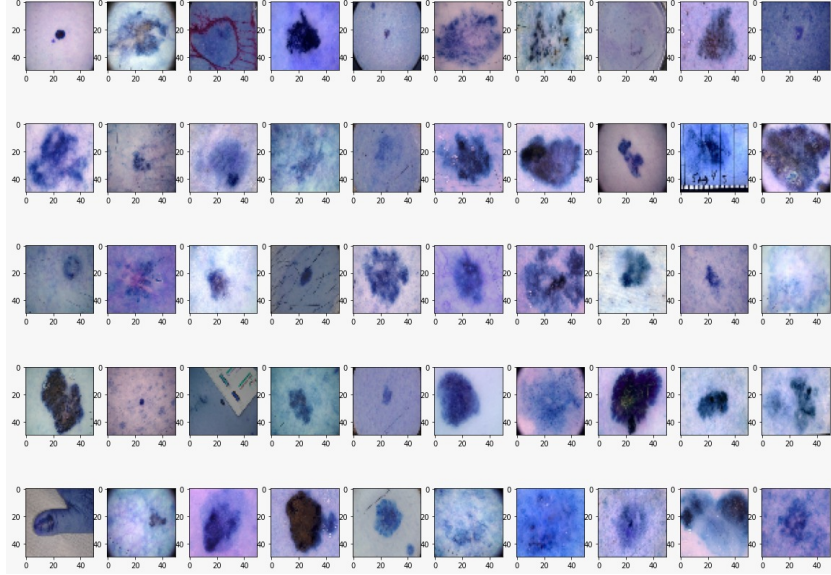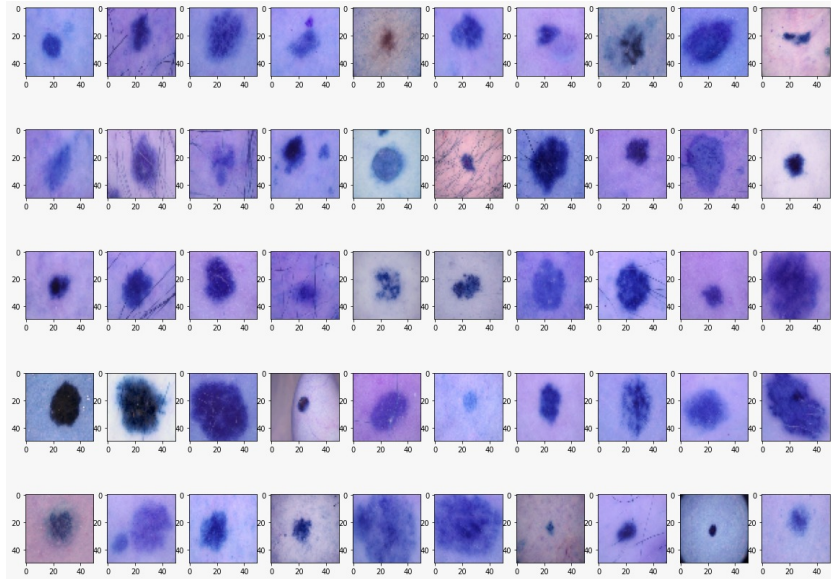
Figure 3: Malignant Skin cancer


Figure 4: Benign Skin cancer

3

2.3   Design

First, we divided the following labels data (0/1 , sick / not sick) that contain 78786 images of breast cancer and 198738 images of health breast into training set and test set. The training set contained randomly 80 percentage of the total images (158990-health and 63029-sick) and 20 percentage of the total images for the test set (39748 - health and 63029-sick). The project contains two part :

1. Check the affect of number of layers on the performance of the net

2. Check if by using transfer learning we can improve the performance of the net

At the first part , we based on exist algorithm [1] for breast cancer detection. In our algorithm we used linear net as shown in fig 5 instead of batch normalization and check the influence of number of layers on the performance.

Figure 5: Neural net- 2 layers
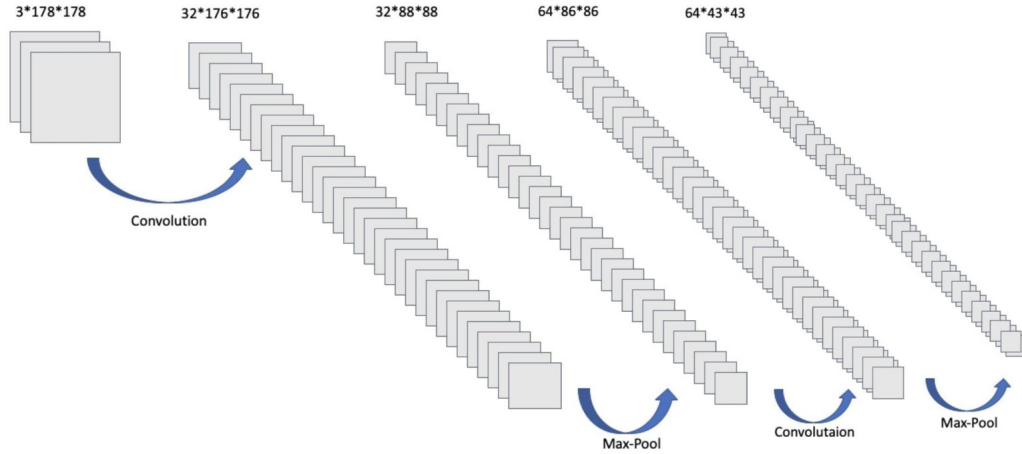
At the second part we used the Pre-trained net of skin cancer and then switch to train it on data set of breast cancer.

3 Experimental results

3.1 Affect of number of layers on the accuracy

We change the number of the layers in order to check what is the affect on the accuracy. The following graph shows the train/test loss Vs. the epochs. As we can see in fig 7, for 3 layers the test loss decrease and then increase. It can hint that we reach to over fit even though that the accuracy of the 3 cases is pretty much the same 87-88 percentage we decided to use the 2 layers net and use it also in the transfer learning.
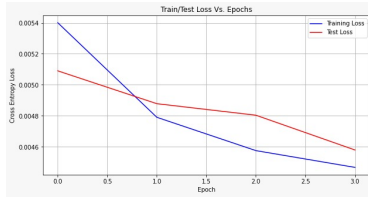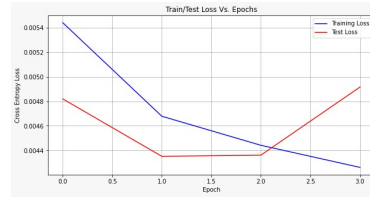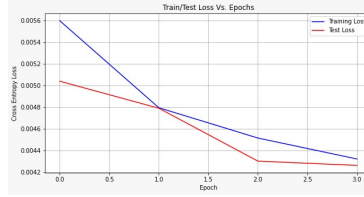

Figure 6: 2 layers

Figure 7: 3 layers

Figure 8: 4 layers

3.2 Transfer learning

By using transfer learning we don't observe any improvement in the accuracy (86 percentage) but we observed improvement in the running time.
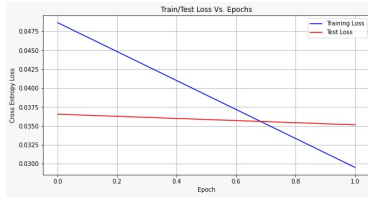

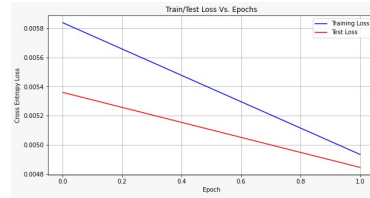Figure 9: Train/test loss Vs. epochs for net

Figure 10: Train/test loss Vs. epochs for pre trained net

4 Discussion

4.1 Challenges

The main challenge was to find data set in a proper size that is not too big and not too small for our project. Eventually we decided to work with 2 data sets that one of them - the breast cancer data set is very big and the second one - the skin cancer data set is too Small. It may lead to incorrect result because we perform the pre train first on the small data set and then did the transfer learning for the big data set.only after discussion with Kfir we understood that we should use it differently.

4.2 Methods

1. In order to get the results that shown before we performed several experiments. Among this experiments we changed the convolution parameters in order to see if the accuracy gets better. Since the data set is very big it takes a lot of computational time and resources and we think that if we had more time to expand the project we could optimize the net much better.

2. During the experiments we noticed that the loss is converge too fast. First we tried 10 epochs and we noticed that it converge for the final loss after 2-3 epochs. So if we have more resources to check the affect of the learning rate and how it improves the performance.

5 Code

`https://colab.research.google.com/drive/1E8uwulk-JuZoIrW7TgsWH9BX-cV6K_Qe?usp=sharing`

References

1. breast cancer implementation - `https://www.kaggle.com/ambarish/breast-histopathology-pytorch`

2. breast cancer data set - `https://www.kaggle.com/paultimothymooney/breast-histopathology-images`

3. skin cancer data set- `https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign`