Breast Cancer Detection Using CNN

Ruggero Barbarossa, Yuhui Bi

Chalmers University of Technology
Electrical Engineering Department

Introduction

As the most common cancer in women, breast cancer is a disease in which cells in the breast grow out of control and it can spread outside the breast through blood vessels and lymph vessels. [1][2]

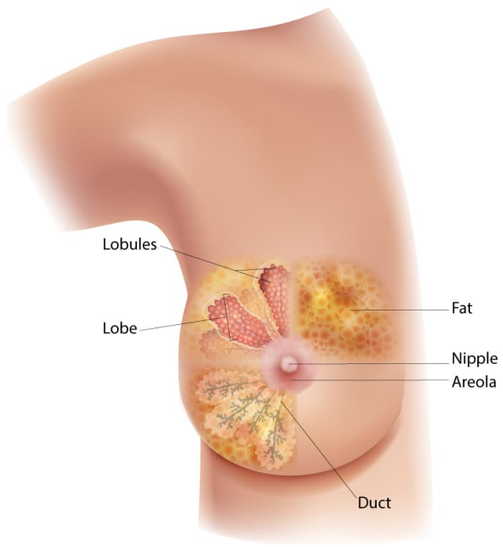


Figure 1: Visualsation of the breast

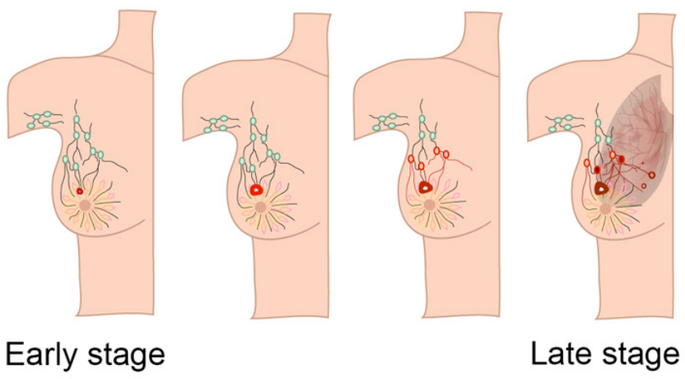


Figure 2: Visualsation of breast cancer

Typical breast cancer:

- Invasive ductal carcinoma (IDC, 80%)
- Invasive lobular carcinoma (ILC, 20 %)

On this project, breast histopathology images of IDC is focused.

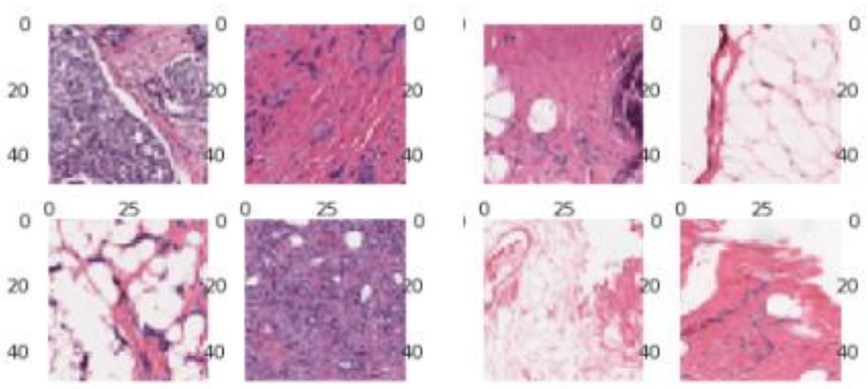


Figure 3: Images from cancer (left) and healthy (right)

Convolutional Neural Network

CNN is a feedforward NN with a deep structure and convolutional computation (optimal for binary classification).

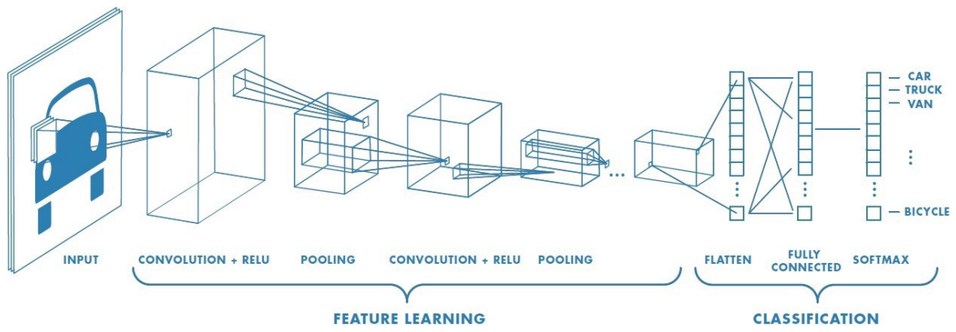


Figure 4: Illustration of CNN

CNN architecture: The input: 50x50 pixels RGB images, and the output: the probability that the image has IDC. Details:

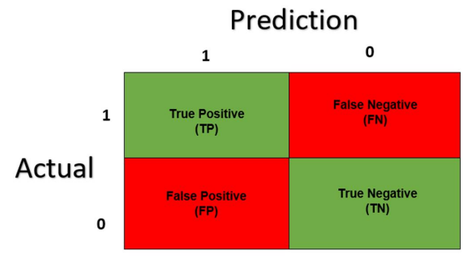
- 2 conv layers
- 2 max-pool layers
- 3 fully-connected layers
  - first 2 fully-connected: ReLU
  - output layer: Sigmoid (produces the probability of cancer)
- Loss function: BCE
- Optimizer: ADAM

Theoretical background

When making classification predictions, 4 outcomes are possible [3]:

- **True positives (TP)**: Prediction $\rightarrow$ Cancer $\|$*Inreal*$\rightarrow$ Cancer.
- **True negatives (TN)**: Prediction $\rightarrow$ Healthy $\|$*Inreal*$\rightarrow$ Healthy.
- **False positives (FP)**: Prediction $\rightarrow$ Cancer $\|$*Inreal*$\rightarrow$ Healthy.
- **False negatives (FN)**: Prediction $\rightarrow$ Healthy $\|$*Inreal*$\rightarrow$ Cancer. Must be minimised

It can be concluded as *Confusion Matrix*



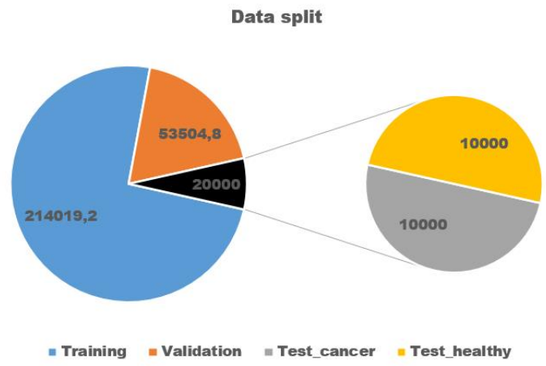
Significant indicators: Accuracy, Recall and **F1-score** (result of harmonic mean of **recall** and **precision**) are:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad Recall = \frac{TP}{TP + FN} \quad F_1 = \frac{\mathrm{TP}}{TP + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})} \tag{1}$$

Results and Conclusions

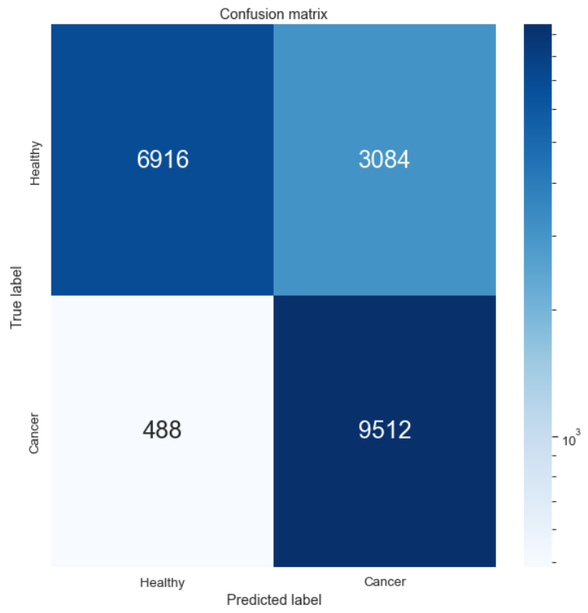
Some optimal procedure:

- **Data balancing**: Most images are healthy $\rightarrow$ unbalanced data.
  - Change class weights (loss function)
  - Upsampling, **downsampling**
- **Data split**: In total, 277524 images (50x50 pixels RGB images)



- **Data augumentation**: Reduce overfitting and boost the performance of the model
  - To flip 50% of images vertically and horizontally
- **Special training algorithm**: Train until converge (min loss + high accuracy)
- **Training procedure**: Minimise FN (maximise recall), keep acc & F1 score high
- **Final model**: correctly classify 95% of samples with cancer, with accuracy of 82%

Results			
Used dataset (threshold)	Accuracy	Recall (cancer)	F1-Score
Imbalanced dataset (0.5)	**0.89**	0.74	0.85
Balanced dataset (0.5)	0.87	0.86	**0.87**
Balanced dataset (0.3)	0.85	0.92	0.85
Test set (0.3)	0.82	**0.95**	0.82



- **Future work**: Apply transfer learning using ImageNet

References

Centers for Disease Control and Prevention. (2022, Spe. 26).*What Is Breast Cancer?*.[Online]. Available from: `https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm`

S.B.Wee. (2018, Oct. 5).*Breast Cancer Screening in Singapore: All about Mammograms*.[Online]. Available from: `https://beta.mountelizabeth.com.sg/healthplus/article/breast-cancer-screening-mammogram`

Barbarrosa, Z. Song, "Efficient Annotations with Deep Active Learning", master dissertation, Dept. Computer Science and Eng., Chalmers Univ. of Tech., Gothenburg, Sweden, 2022.

rugbar@student.chalmers.se; yuhuib@student.chalmers.se