

Capstone Proposal

Breast cancer classification using Transfer Learning

Murali Mallikarjuna rao Perumalla
perumallamurali857@gmail.com



School of Artificial Intelligence
Machine learning engineer nanodegree

I. INTRODUCTION

Breast cancer is most commonly occurring cancer in the women and it's the second most common cancer overall. According to the 2018 statistics, there were over 2million cases all over the world. Belgium and Luxembourg have the highest rate of cancer. Early detection of breast cancer has reduced many deaths. Earlier CAD systems used to be second opinion for radiologists and clinicians. Machine learning and deep learning has brought tremendous changes in medical diagnosis and imagining. Early detection and classification of cancer helps specialists to diagnose accurately and reduces the patient re-visit rate.

II. PROBLEM STATEMENT

It is known that Deep learning needs very huge amount of data and high computing resources. In case of breast cancer almost all datasets are small. And dataset is imbalanced. Small datasets are hard to converge. Imbalance in datasets leads to the overfitting, less test accuracy.

III. DATASET AND INPUTS

In this project I am using BreakHis breast cancer dataset. The Breast Cancer Histopathological Image Classification (BreakHis) is composed of 9,109 microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To date, it contains 2,480 benign and 5,429 malignant samples (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format). This database has been built in collaboration with the

P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil (<http://www.prevencaoediagnose.com.br>). We believe that researchers will find this database a useful tool since it makes future benchmarking and evaluation possible.

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

The dataset currently contains four histological distinct types of benign breast tumors:

- adenosis (A),
- fibroadenoma (F),
- phyllodes tumor (PT),
- tubular adenoma (TA);

four malignant tumors (breast cancer):

- carcinoma (DC),
- lobular carcinoma (LC),
- mucinous carcinoma (MC)
- papillary carcinoma (PC).

IV. SOLUTION STATEMENT

Solution is applying transfer learning models. To deal with the imbalance classes I am using pytorch sampler known as weighted random sampler. In this project I will use various transfer learning models and use model with best test accuracy.

V. BENCHMARK MODEL

In most of the recent research works perform binary classification i.e. cancer is benign or malignant. In this project I will classify the cancer into 8 different cancer types . without transfer learning , model achieved train accuracy, train loss, validation as 60% , 0.5678, 0.9654 respectively . we can also see that model overfits.

VI. EVALUATION METRICS

To evaluate the model, I am using sensitivity and specificity as evaluation metrics.

VII. PROJECT DESIGN

- **Programing Language:** python3.6
- **Libraries:** Torch, Torchvision, matplotlib, OS, Numpy , Glob
- **Workflow:**
 - Build the Dataset into train, test and validation folders
 - Load the dataset using pytorch Data loaders and Image folders
 - Download the pretrained models
 - Perform training
 - Hyperparameter tuning
 - Test model and predict
 - Plot the predictions