



A comparison of the efficiency of MRI and Mammography in breast cancer detection, using Convolutional Neural Networks

Keneilwe Mokoka

School of Computer Science and Applied Mathematics

Research Question

Can MRI detect breast cancer better than mammography?

Research Hypothesis

- H1: MRI is more efficient than mammography examination in detecting breast cancer.
- H0: MRI is not as efficient as mammography examination in detecting breast cancer.

Research Aim

The aim of this research is to compare the efficiency of MRI and Mammography in breast cancer detection using Convolutional Neural Networks

Research Objective

The objectives of this research are:

- Build a CNN in Python to classify tumors from MRI images and mammograms as benign or malignant.
- Compare the performance of the CNN on the MRI dataset and mammograms based on classification accuracy and confusion matrix (to determine true/false positives as well as true/false negatives)

Research design

A CNN will be implemented in Python. The MRI data set will be split into two: the training and the validation data. The network will be trained with the data set aside for training and the validation data will be used to check the accuracy of the network. Similarly, the mammography data set will be split into two: the training and the validation data. A comparison of the two will be done, to conclude whether the CNN performed better for the MRI data set, compared to the mammography data set. The network will also be evaluated on its true/false positive rates as well as true/false negative rates for both the MRI and mammography data sets.

Data

The data sets will be sourced from the Cancer Imaging Archive (CIA). Both the MRI and mammography data sets are in DICOM format. The breast MRI data set consists of a collection of 922 biopsy-confirmed invasive breast cancer [2]. The data set is ideal since the locations of the lesions have been located on the images, and features such as the size, shape and texture of the tumor have been identified. An additional pre-processing activity will include standardizing the images into 50*50*1 pixel images and normalizing the pixels into values that are between 0 and 1.

Data

The mammography data set contains region of interest segmentation as well as bounding boxes [1]. This will in turn reduce the efforts of pre-processing the data. The data set consists of normal, benign and malignant cases. It has already been split into two: the training data set and testing data set.

Methods

Python will be used to resize the images into the same size as an attempt to standardize the data. The data will be normalized between 0 and 1. To implement the CNN, the following Python libraries will be used:

- Keras
- numpy
- matplotlib
- scipy.

Analysis

The following will be used to measure the performance of the CNN:

- Classification accuracy
- Confusion Matrix (False/True Positives, True/False Negatives)

Acknowledgements

I would like to extend my gratitude to Dr Michael Olusanya for providing ideas and feedback regarding my research.

Bibliography

- [1] Lee, Rebecca Sawyer et al (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. In *Scientific data* 4.1, pp.1-9.
- [2] Smith, K. CBIS-DDSM. Retrieved from: <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM225166295e40bd1f79d64f04b40cac57ceca9272>