

Heinz 95-845: Project Proposal

Tanmaya Tripathi

Heinz College

Carnegie Mellon University

Pittsburgh, PA, United States

TTRIPATH@ADDRESS.EDU

Editors:

1. Proposal Details (10 points)

1.1 Proposed Analysis

The study proposes researching and developing methodologies to identify tumours in MRI images. The analysis will use TensorFlow open source software library on an existing set of MRI images. The likely outcome of the study would be that given a specific AX T1 set of images for a patient the tool should be able to identify if any of the images has a presence or hint of tumour.

1.2 Importance of the analysis

Cancer currently, is one of the leading causes of death in the medical field. As of 2012 cancer was responsible for 14.6% of overall human deaths. Currently, oncologists are struggling to keep up with the cancer related cases. This analysis is aimed at finding methods that will identify the tumours and will in turn make it less cumbersome for the doctors while dealing with cancer related cases.

References-

<https://en.wikipedia.org/wiki/Cancer>

1.3 Contribution of Analysis to existing work

The current research/data on cancer is research focussed where the end results are available only for doctors, hospitals etc. Furthermore, the technology is novice and has not found extensive use in third world countries due to issues like cost and lack of technology development. The analysis aims on making an open source ended effort which would hopefully be used to propel further research and development in open source solutions.

1.4 Description of data

The data for this analysis will be taken from Cancer Imaging Archive. The following are the characteristics of the data and analysis-

Y outcome(s) - given an MRI image, analyzing and predicting whether the image contains tumour

U treatment - the model will be developed using deep learning algorithms

V covariates - at this stage there are no covariates since the model will be developed using an existing set of MRI images

W population - the training set contains 262 patients

References-

<http://www.cancerimagingarchive.net/>

1.5 What evaluation measures are appropriate for the analysis? Which measures will you use?

As stated above, the analysis will use deep learning algorithms in order to process the MRI images and identify tumours if present.

1.6 Study design and ML algorithms to be used

The analysis aims to use deep learning algorithms as part of TensorFlow package in order to process the MRI images and identify the tumours. Currently, the data consists images from scans of 262 patients. The total size of the data set is approximately 73.5 GB. The initial step would be to pre process the images and extract only the AX T1 images for further analysis. This would be followed by processing and analyzing the reduced data set in python using the TensorFlow package.

1.7 Limitations of the study

The study currently has the following limitations-

1. Images in the data set are not a part of research study. They have been collected during natural data collection process. Hence, there are significant chances of errors or blurred images in the dataset.