Investigating effectiveness of various CNN architectures for medical imaging with a focus on Alzheimer's studies

Name: Amelia Kate Moseley **Student Number: 1806461** **Course: BSc (Hons) Computing (Software Engineering)**

Project Supervisor: Gordon Dickers

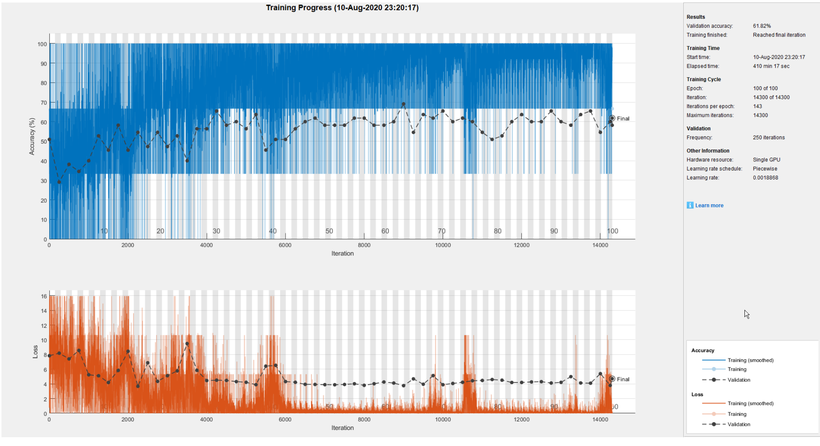
1 - Introduction

Automatic diagnosis for biomedical imaging tends to use the traditional machine learning models and methods which require people to build manually designed feature sets based on the raw data, whereas Deep Neural Networks (DNN, the above mentioned CNNs being examples of DNNs) extract features automatically. Attempts have been made in the 90s to utilize CNNs for biomedical imaging but did not prove reliable enough to justify replacing traditional machine learning methods and manual feature extraction. As larger datasets become more readily available and accessible, and the processing power of GPUs increases, one must consider the use of CNNs for automatic diagnosis and other biomedical classification. For the purposes of this paper a comparison is to be made between to methods of classification using CNNs: 2D image classification based on slices of MRI volumes, like most methods of biomedical imaging in terms of the use of slice classification, but using CNNs instead of manual feature extraction; and the second method is full 3D volume classification, classifying the whole volume based on given labels. The comparison in this paper is for the classification of Alzheimer's Disease (AD); identifying between the three main categories of AD classification: AD, Mild Cognitive Impairment (MCI) – the prodromal stage of AD, and Cognitively Normal controls (CN).

2 - Project Aims

The main aims for this project were to explore different possible architectures to conduct the training with, and to try to achieve satisfactory accuracy, sensitivity, and specificity in classifying between: AD, MCI, and CN.

The goal was to try and find 2 to 4 suitable architectures to compare how well they performed, with the hope being that performance metrics of ~75% would be reached. The architectures found to be used were: 3D U-Net (already used for biomedical imaging with segmentation), ResNet 101 (a successful architecture for 2D classification), and a paper which discussed the potential use of a hybrid LSTM-CNN (Long Short-term Memory). Of these 3D-Unet and ResNet were used as the other was unable to be implemented.



3 - Methods

Training was conducted in MATLAB using the Deep Learning Toolbox, with basic scripts used to process the data as it gets read it (namely a resize function whose handle gets passed to the built-in functions when they read the data from the datastore. Due to the high memory requirements of training using 3D data, an ex-server accelerator card was used for a large portion of the training attempts; to get this working it required a custom cooling solution. Data used for the project was acquired from the Alzheimer's Disease Neuroimaging Initiative after permission for use was requested.

4 - Results

The network training did not yield any suitably accurate models. The initial training attempt was carried out on the GPU using the standard 3D U-Net architecture and achieved the second highest accuracy, sensitivity, and specificity with each of those metrics sitting at ~61%. Subsequent training yielded far lower metrics at around 33%.When the classification method was switched to two possible classes, the metrics were an average of 55% for each training attempt; this was the same for each network architecture used, this being a binary classification puts the 55% at just over chance and therefore on par with the ~33.21 from the three possible classes.

5-Discussion

Looking at the way the project unfolded in retrospect, there are plenty of are-as for improvement, lessons that have been learned, and potential ways to carry out a derivative of the project in the future. From the start it should have been obvious that it was too ambitious a project to carry out at this level of study; the data being used and the decision to carry out the training using the full volumetric MRI was not a good idea. There is potential merit in full volume classification, it is just that that specifically should have been the focus of the project, removing the superfluous nature of the medical aspect of things; too many aspects of interest were being brought together for a project of this level. The implementation of the hybrid LSTM-CNN could have been an interesting way to look at using full 3D data as inputs to an architecture, the potential increase in accuracy gained from using the short-term memory to build connections between the layers of the 3D volume.

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David