Cardiff University
Computer Science and Informatics

CM3202 - Individual Project

AI to Detect Early Signs of Alzheimer's Disease

Initial Plan

Author:
Jacob Winkworth

Supervisor:
Xianfang Sun

Contents

1 Ethics

The data for this project has been provided by the Alzheimer's Disease Neuroimaging Initiative. The ADNI is a multicentre research program with the aim to develop clinical, imaging, genetic, and biological biomarkers for the early detection and tracking of Alzheimer's disease.

The three main objectives of the study are:

1. To detect AD at the earliest possible stage and identify ways to track the disease's progression with biomarkers.
2. To support advances in AD intervention, prevention, and treatment through the application of new diagnostic methods at the earliest possible stages.
3. To continually administer ADNI's innovative data-access policy, which provides all data without embargo to all scientists in the world.

All ADNI data is shared without embargo through the LONI Image and Data archive (IDA), a secure research data repository [1]. An application for permitted access to these archives has been signed and submitted by researchers at Cardiff University, of which the login details have been shared for the intents and purposes of this project. During this application process, the applicant must agree to the Data Use Agreement as provided by the ADNI. Thus, these criteria will be firmly adhered to. This data set will be analysed via a password-protected PC at the facilities of Cardiff University with GPU-enabled PyTorch (CUDA).

The ADNI data set includes the following types of data:

- Clinical (demographics, clinical assessments, and cognitive assessments).
- Genetic (Illumina SNP Genotyping).
- MRI image files.
- PET image files.

The main samples of interest for this project are the MRI image files, but further development of the software could lead to the use of multiple data types. All samples are anonymised, and no attempt to establish the identity of or contact any subject will be made, as part of the ADNI Data Use Agreement [2].

2 Project Description

With the average age of the global population on the rise, Alzheimer's disease is soon expected to reach epidemic proportions [3]. Advancements in neuroimaging are currently leading the field in the diagnosis of the condition, but the earlier these signs are detected the more effective the treatment for the reduction of symptoms; a syndrome more commonly known as Dementia. This project aims to develop and train a deep learning model with the ability to distinguish Alzheimer's magnetic resonance imaging (MRI) from normal, healthy control data. The system will later progress to predict whether a patient will develop Alzheimer's after first being presented with mild cognitive impairment (MCI).

An MRI is a common imaging technique used predominantly in the medical industry for establishing a 3D portrait of anatomy in a digitized form [4]. This assists in the diagnosis of patients suffering from a multitude of various mental and physiological health defects. Such defects caused by Alzheimer's disease include abnormalities associated with mild cognitive impairment (MCI) and the decrease in size of multiple regions within the brain. These regions include the hippocampus (temporal lobe) (the forming of new memories), the frontal lobe (intelligence, judgement, behaviour) and the parietal lobe (language) [5].

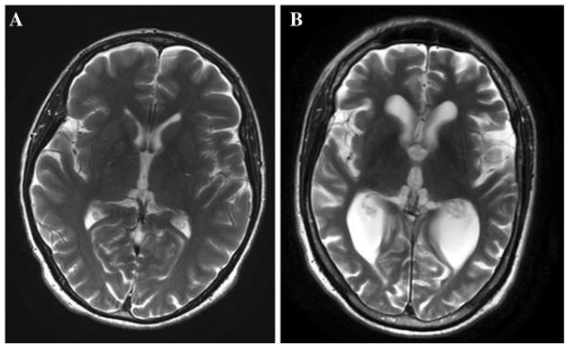

Figure 1: Alzheimer's (A) vs Normal (B) MRI brain scans [6]

Moreover, this project will follow a predictive modelling approach to deep learning. This can be further divided into classification predictive modelling; the task of approximating a mapping function (f) from input variables (x) to discrete output variables (y) [7]. Training a deep-learning model under this process will allow for the identification of subjects with an elevated risk of subsequent Alzheimer's, given their current cognitive impairment. The deep learning model will employ a convolutional neural network (CNN). In general, convolutional neural networks are compiled of two main stages: feature learning and classification. The feature learning stage holds the major distinction from traditional networks. During this phase, a high-volume of training data undergoes a series of kernel convolutions to isolate scale- and shift-invariant features. As the layers progress, the features become more complex and distinguished. These kernel convolutions are otherwise known as filters and

allow the CNN to capture spatial and temporal dependencies in an image. This provides a better fitting to the dataset, as the number of network parameters is dramatically reduced, without the loss of critical attributes. After a series of convolution and pooling stages, the final output is flattened into a column vector and fed through a multi-level perceptron for classification [8].

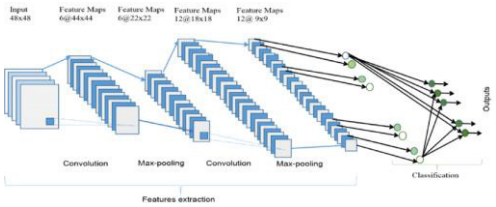

Figure 2: Example CNN [8]

There are various CNN architectures openly available to utilize, notably VGGNet, LeNet, AlexNet and GoogLeNet. These are set in place and shared with the community to allow for further development of algorithms based upon well-tested and robust methods. This project will follow one of these architectures. More precisely, the proposed deep-learning model will employ a method of transfer learning, whereby several layers from a trained model will be reused, retrained, and specialised for the intent of this project.

As per the predictive modelling approach, there will be three disjoint subsets compiled from the ADNI database for training and validating the network. The first of which will be a high-volume collection, classified further into a subset depicting subjects with confirmed Alzheimer's, and a subset of normal healthy controls. This is the data set the weights and biases of the neural network will adjust to fit. It's important, however, to avoid overfitting the network. This occurs when a model achieves near-perfect results with a particular set of data but struggles to generalize, resulting in poor performance when presented with unseen data. Next, the validation set will be parsed for frequent evaluation of the model, enabling the fine-tuning of hyperparameters manually. Finally, the test set will be used to evaluate the success of the final model. The data provided by the ADNI includes a comprehensive demographic background of each subject. Thus, data sets can be further categorised into age groups, gender and genetic backgrounds where necessary.

To surmise, the overarching aim of this project is to develop a deep learning-based pipeline using a convolutional neural network to distinguish MRI scans displaying subjects with Alzheimer's from those without. With precise training and configuration of the network, an increase in sensitivity should enable the model to classify the scans of subjects in much earlier stages of the disease. Eventually, detection sensitivity should reach a point at which the network correctly predicts and classifies inputs that display minimal signs of the disease, namely mild cognitive impairment.

3 Project Aims & Objectives

The aims of this project are tailored towards the development of a convolutional neural network to predict whether a patient will develop Alzheimer's after first being presented with MCI. The system will be given raw MRI inputs and focus upon structural changes within the brain that represent the deterioration of tissue caused by the disease. The aims are outlined as follows:

Aims

- Develop a CNN to detect outstanding signs of Alzheimer's disease from subjects where the disease has matured, trained from the ADNI data set.
- Develop a CNN to detect early signs of Alzheimer's where the subject displays mild cognitive impairment, trained from the ADNI data set.

The following objectives outline how the above aims will be achieved:

Objectives

- Thoroughly research classification predictive models to determine several formats to evaluate. Such may include:
  - Logistic regression.
  - Decision trees.
  - Random forest.
  - Naïve Bayes.
- Thoroughly research convolutional neural network architectures and determine a suitable format to follow. Such may include:
  - VGGNet
  - LeNet
  - AlexNet
  - GoogLeNet
- Develop multiple instances of a convolutional neural network under various classification models to detect Alzheimer's in subjects where the disease has matured:
  - Evaluate the success of each model.
  - Select a best-fitting model.
- Use best-fitting model for development of a convolutional neural network to predict the development of Alzheimer's in subjects displaying signs of mild cognitive impairment.

4 Work Plan

Deliverables

The following files are to be submitted by the deadline:

- Final report
- All source code
- Supporting documents
- Graphs, tables and other project-related data visualisation

Research

Such a task will require a detailed understanding of several topics within artificial intelligence and deep learning. These include predictive modelling and convolutional neural networks, along with their underlying aspects. Thorough research will be carried out to ensure this knowledge is gained prior to the implementation of the network. Building on this, adeptness in proposed frameworks (e.g. PyTorch) is also required, which will be sought after via relevant tutorials and documentation sourced online.

Implementation

Following a suitable level of research, implementation will begin with the development of a CNN to distinguish subjects with matured Alzheimer's disease from those without. Different classification models will be experimented with and evaluated to determine one that best-fits the project. From there, the model will be developed further into a CNN that can detect early signs of Alzheimer's where the subject displays mild cognitive impairment

Supervisor meetings

Several meetings with the project supervisor will be scheduled over the period of this project. These will help with guidance on the subject and ensure work-rate is consistent.

Milestones

With the allowance of a one-week extension to the initial plan deadline following extenuating circumstances, this is the proposed workplan for the project:

- **Week 1** 03/02/2020
  - Download and install relevant software and libraries to personal computer: PyTorch, Jupyter (or similar Python IDE) etc.
  - Initial meeting with project supervisor (along with a member of the university's medical research team to aid with the ADNI data).
  - Develop initial plan.

- **Week 2** 10/02/2020
  - Submit initial plan.
  - Research topics relevant to the project.
  - Follow online tutorials to develop a simple neural network to aid with the understanding of PyTorch (the SentDex Youtube channel is an excellent source for this).

- **Week 3 - 6** 17/02/2020
  - Develop a convolutional neural network to distinguish MRI scans showing the presence of matured Alzheimer's in a subject from normal, healthy control data. Two or more classification models should be tested during this process.

- **Week 7** 16/03/2020
  - Evaluate the success of each classification model using suitable methods of statistical analysis.
  - Decide upon best model to continue development with.
  - Begin the development of a convolutional neural network to predict whether a subject will develop Alzheimer's given symptoms of mild cognitive impairment.

- **Week 8 - 12** 23/03/2020
  - Develop a convolutional neural network to predict whether a subject will develop Alzheimer's given symptoms of mild cognitive impairment.

- **Week 13** 20/04/2020
  - Evaluate success of full implementation using suitable methods of statistical analysis.
  - Begin writing report.

- **Week 14 - 15** 27/04/2020
  - Write and submit report.

References

[1] Adni.loni.usc.edu. (n.d.). *ADNI | Alzheimer's Disease Neuroimaging Initiative*. [online] Available at: http://adni.loni.usc.edu/ [Accessed 03 Feb. 2020].

[2] ADNI Data Use Agreement. (2016). [ebook] ADNI. Available at: http://adni.loni.usc.edu/wp-content/themes/freshnews-dev-v2/documents/policy/ADNI_Data_Use_Agreement.pdf [Accessed 03 Feb. 2020].

[3] Shapiro, S., Shapiro, D. and Hamby, C. (n.d.). *Alzheimer's disease: an emerging affliction of the ageing population*. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2931467 [Accessed 04 Feb. 2020].

[4] Lewis, T. (2017). *What is an MRI (Magnetic Resonance Imaging)?* [online] livescience.com. Available at: https://www.livescience.com/39074-what-is-an-mri.html [Accessed 04 Feb. 2020].

[5] National Institute on Aging. (n.d.). *What Happens to the Brain in Alzheimer's Disease?* [online] Available at: https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease [Accessed 05 Feb. 2020].

[6] Cambria, M. (n.d.). [online] Available at: https://www.researchgate.net/figure/Brain-MRI-Axial-T2-images-shows-cerebral-atrophy-in-patient-with-Alzheimers-disease-A_fig1_260115752 [Accessed 05 Feb. 2020].

[7] Brownlee, J. (2017). *Difference Between Classification and Regression in Machine Learning*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/ [Accessed 06 Feb. 2020].

[8] Saha, S. (2018). *A Comprehensive Guide to Convolutional Neural Networks*. [online] towardsdatascience. Available at: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [Accessed 06 Feb. 2020].