

Bachelor Thesis - Report Defense

Alzheimer's Disease Detection

Ichim Ștefan

22 May 2024

Contents

1	Challenges	2
2	Strengths	2
3	Weaknesses	2
4	Opportunities	3
5	Threats	3
6	Impact	3
7	Acknowledgement	4

1 Challenges

Throughout these weeks of research, there have definitely been an arsenal of challenges, especially those that tackled my data managing techniques. Up until now, most of my machine learning experiments had been using known image formats(jpeg, png), text inputs or real numbers, but this particular work had me discover another format, the NiFTi(.nii) one.

To begin with, this NiFTi format has the purpose of protecting very important voxels of 3D MRI scans, and as such even visualizing such files, or creating them also requires additional steps.

Another challenge was adapting with how the data was given by the ADNI. Their filtering website is beyond outdated, in the sense that it hasn't received updates for a decade. Files which you think would be valid would turn out invalid, missing segmentation mask counterparts or labels, or missing at all when a .csv which establishes some boundaries of control over the data would mention it should exist.

Furthermore, it was the first time i had run out of memory on the GPU while training a model. This led to further considerations on what to send to the GPU device, and not think blindly it can handle anything as it gets fed.

However, the greatest challenge was the inability of using proper preprocessing tools which were popular among the state-of-the-arts.

2 Strengths

The main reason for why I chose this topic was to learn two major components: how the Alzheimer's Disease manifests neurologically speaking, and how to tackle such a problem of Computer Science. My main interest was how these two combined together to solve the problem, firstly how the Alzheimer's Disease distinguishes itself among other neurological diseases, how these areas are then observed and stored, and finally how Computer Science's branch of Machine Learning intervenes in this. What do researchers highlight for these algorithms to learn and why. And I consider to have at least grasped some of their intentions and reasoning behind why certain input times become valued above others in which cases.

Besides that, my python capacities were improved, after using more of PyTorch throughout the development of these experiments. By creating my own model with each layer from scratch, and bringing adaptations to it after having read about them in different articles of the state-of-the-arts, which I considered worthy of adding to my own model.

I also understood more how the CUDA environment works, and not just use it blindly and expect it to always do the right job.

3 Weaknesses

The greatest issues met during this study was the lack of preprocessing available. While the ADNI provides some segmentation masks for a few of their scans, researchers in fact avoid using the provided masks at all, calling on to developed softwares which come with a variety of functionalities. Among their most important, the higher quality segmentation is one of them, which properly realizes

the skull-stripping step, and, secondly, the image registration, which represents the ability of bounding each scan to the same coordinates. Due to the lack of these two great aiding tools, I was only left with the primitive segmentation masks offered by the official ADNI Initiative.

Another weakness includes the unexpected size of the dataset. Since the images came with the original scan and the segmentation mask, and then the mask had to be applied, that would mean 3 images would need to be stored for one entry. Thus it lead to the size of 381GB for 400 patients after all the masks were applied.

With all this mind going further, I was optimistic to achieve an accuracy result around 65% and 70%, but i first started out at 56%, and could only improve up to 61%.

4 Opportunities

As mentioned in the previous section, improvements can be made in a variety of ways.

Firstly, an obvious step would be increasing the dataset size by expanding to phases that followed the ADNI1. A few include: ADNI-GO, ADNI2, and recently ADNI3 and ADNI4. This would generate even more subjects with various visits, or new patients which joined in the later phases.

Secondly, the Harvard's FreeSurfer software with its complex algorithms could be used for both segmentation and image coordination fixing. However it was only available on Linux and MacOS, so its functionalities couldn't be applied.

As a last improvement idea, a GPU with higher memory capacity could be used which could allow for multiple slices to be sent altogether, not a specific one for each scan.

5 Threats

The main risks involved was actually while gathering the necessary data. ADNI does not provide a good enough guide of how the images are stored, and finding out the specific ones which also had higher chance of being accompanied by their respective segmentation mask at some point didn't seem feasible. And having to use as input images the raw scans where both the brain and the skull would have to go together in the classification algorithm would lead to even more erroneous data. However with further digging, forums searching and finding better instructions, some images with their respective masks could be found, downloaded and used.

6 Impact

My work impact can be seen by how important the developed preprocessing softwares actually are in this particular field. It is under no circumstance possible to ask patients to take the MRI scans while sitting in the same position, due to either body size differences, or other sitting issues. And as such it is vital to

apply the complex algorithms which are popular throughout the study in this field.

With that being said, an accuracy result of 62% was still reached using only 487 scans from 300 different subjects. Thus with an expanding of this amount of patients, better results could be obtained.

7 Acknowledgement

This work is the result of my own activity, and I confirm I have neither given, nor received unauthorized assistance for this work.

I declare that I did not use generative AI or automated tools in the creation of content or drafting of this document.