Capstone Proposal

Kristin Levine – Spring 2022

Using ensembled pre-trained CNN networks to classify MRI images of patients with Alzheimer's Disease

Advisor: Amir Jafari

The George Washington University, Washington DC

Data Science Program

1 - Objective:

The goal of this project is to use MRI images to help decide if a patient is cognitively normal (CN), has mild cognitive impairment (MCI), or has Alzheimer's disease (AD).

2 - Dataset:

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a public-private collaboration studying Alzheimer's disease. The initiative has been around since 2004 [1] and collects clinical, cognitive, imaging, biomarker, and genetic data to share with researchers.

3 - Rationale:

Alzheimer's disease is the most common dementia, yet it remains difficult to diagnose. PET scans to measure amyloid protein build-up and cerebrospinal fluid (CSF) biomarkers can help to provide an accurate diagnosis, however they are expensive and invasive tests. MRIs are much more accessible.

However, MRIs are difficult to read and depend on an experienced neurologist to interpret them. A dependable and accurate ML learning model has the potential to help a general practitioner in a small town to read a MRI as accurately as a specialist at a large research hospital. There has been progress in developing ML models for other conditions like breast cancer. [2]

Being able to differentiate mild cognitive impairment from normal cognition and Alzheimer's disease is also important for clinical trials, as potential treatments are more likely to be effective before the condition progresses to AD.

4 - Background:

In the past, neurological ML models mainly used MRIs for feature extraction, i.e. to get a measurement for a part of the brain, such as the size of the hippocampus. While this has proved helpful, there is significant domain knowledge required in order to understand which parts of the brain to target. Extracting features is challenging; we also can't be sure that we are picking out the most helpful features.

CNNs offer a different approach of using computer vision to look at the image as a whole. There is less domain knowledge needed as the computer decides which are the relevant features.

However, using a CNN network is not without its own challenges. CNN networks work best when you have lots of data; however, in dementia research there is a lack of labeled images. While I've read about some interesting approaches, including using unlabeled data to help train a network [3], in the interest of simplicity, I think for this project I should stick to labeled datasets.

If you don't have a ton of images, one way to get around that drawback is to use transfer learning, i.e. using pretrained networks as a starting point for your model. While there has been some discussion about whether it is useful to use networks pre-trained on ImageNet with medical data [4], I'm going to investigate pretrained models for this project. There are plenty of other studies that have used these pretrained models to look at Alzheimer's imaging data. [5] [6] [7]

Unlike X-rays which I have worked with in the past, MRIs are, of course, 3D images. While 3D CNN networks are a new area of research [8] and sound very cool, in the interest of creating a project I can actually get done and doesn't require extremely large amounts of computing power, I'm going to stick to 2D for this project. This 2D approach has been used by other researchers. [5] [9]

5 - Approach:

I plan on approaching this capstone through several steps.

1. Familiarize myself with MRI file formats. These are typically DICON or NIFTI files. I'm currently exploring some online tutorials about working with these types of files.

2. Familiarize myself with the dataset. A colleague at my job is currently working on a feature extraction project with this data. I think he will be a good resource in helping me to understand the organization of the dataset.

3. Extract 2D slices from each MRI in the ADNI dataset and preprocess the images. It will be interesting to decide how to do this. Some studies have taken all the slices, some have

selected a small range of slices,[9] and others took what they considered the 32 most informative slices. [10]

I'm thinking about taking a middle approach: thirty slices from each patient. I also want to make sure I get different orientations from the slices, perhaps 10 axial, 10 coronal, 10 sagittal views. My views on this approach may change as I do more research. See examples of slices from the dataset below:
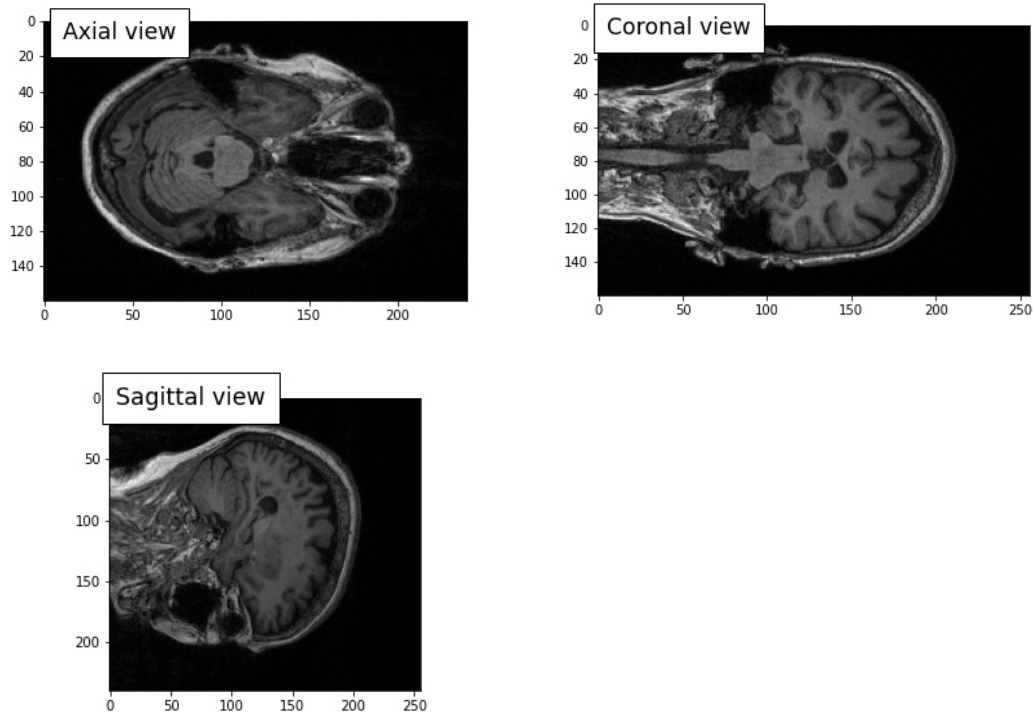
For some patients, there is more than one MRI session – I'll have to decide which session to use. Some patients shift in their category over the course of the study - from CN to MCI or MCI to AD.
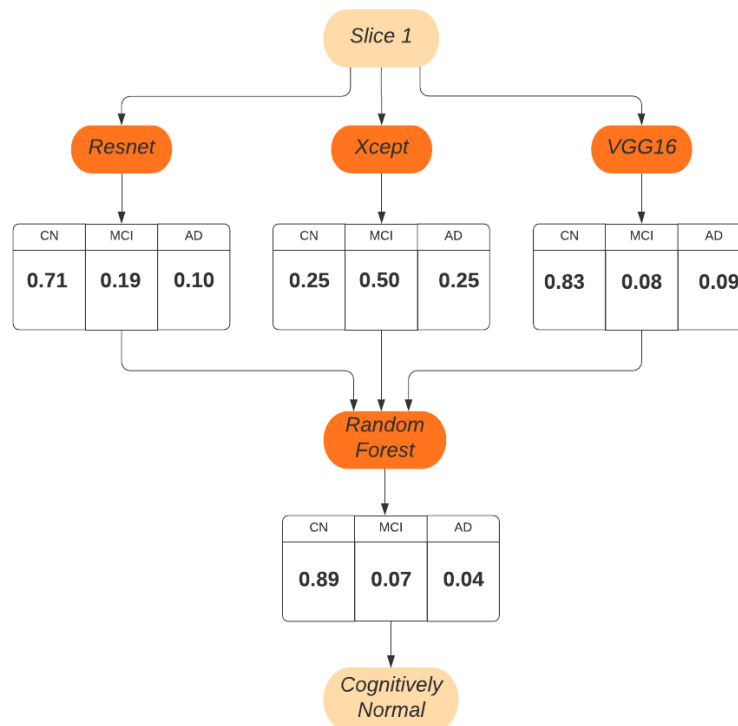
4. Divide the ADNI images into training, validation, and testing sets. Train the model on the ensembled CNN network we used on chest X-rays in ML2. I've also used this same model on brain tumor images in my Cloud Computing class.

    This model uses ResNet, Xception, and VGG16 to get a probability for each category, in this case CN, MCI, or AD. The probabilities from each pre-trained model are then combined in a new dataframe and a Random Forest is used to get a final prediction.

    See example of architecture below with hypothetical values:

Ensemble CNN Architecture
Kristin Levine | January 17, 2022

Slice 1

Resnet

CN	MCI	AD
0.71	0.19	0.10

Xcept

CN	MCI	AD
0.25	0.50	0.25

VGG16

CN	MCI	AD
0.83	0.08	0.09

Random Forest

CN	MCI	AD
0.89	0.07	0.04

Cognitively Normal

I'm going to try all the typical parameters adjustments, such as freezing/unfreezing final layers, changing batch sizes, LR, etc.

Since we are interested in an AD prediction per patient (and not per slice) I will then combine all 30 sets of probability scores (if we use 30 slices) in another RF and get a prediction of CN, MCI, or AD per patient.

I'm not sure yet if it would be better to take all the raw data and run one large RF (9 probabilities per image, 30 images = 270 columns in DF) or if it would be better to get the probability for each image (3 x 30 = 90 columns) and then run a second RF to create a final prediction.

5. (Optional) If time allows, repeat on a different dataset to see how well the model generalizes.

6. Create a website like this to display results. This was a project for the Cloud Computing class looking at tumors in brain MRIs. Instead of only 3 slides per patient, all the slices

from a particular patient would be displayed, so that the user would be able to see how confident the model was in its predicted results.

6 - Timeline:

This is a rough timeline for this project:

- (2 Weeks) Learn everything I can about MRI images
- (2 Weeks) Preprocessing images
- (2 Weeks) Extracting slices
- (3 Weeks) Training ensembled CNN models
- (1 Week) Compile Results
- (2 weeks) Create website to visualize results
- (2 Week) Writing up paper and submission
- (1 Week) Final Presentation

While I would like to test any models I come up with on a second dataset, if I run out of time during the semester that would easily be something I could do as a follow up project.

7 - Similar Projects/How mine will be different:

Most of the similar studies I found used only one pretrained network to classify the images, although they sometimes compared the results from more than one. Some of the pre-trained CNNs used were: VGG-19 [5], ResNext101 [6], GoogLeNet [11], and Inception [12].

Fang et al's 2020 study [9] is the most similar one I've come across to what I'm proposing. They also extracted 2D slices from the MRI and PET images, and then used three different pre-trained networks (GoogLeNet, ResNet, and DenseNet) to get a probabilistic score for each image. Finally, they used a Decision Tree to combine the probabilistic scores from each CNN and get their final prediction.

My proposal is different in the following ways:

	Fang et al	My Proposal
CNN models used	GoogLeNet, ResNet, and DenseNet	ResNet, Xception, and VGG16
Image type used	MRI and PET	MRI
Number of slides (per patient)	16 slices per MRI	30 slices per MRI
View used	Coronal	Coronal, axial, and sagittal

Algorithm to combine probabilities	Decision Tree	Random Forest
Classification type	Binary: CD vs. AD	Multi: CN, MCI, AD

8 - Expected Number of Students:

This would be an independent project; in addition to the support of the Capstone class, I would also use colleagues from work as resources.

9 - Possible Issues:

Working with MRI images

I've never worked with MRI images before, except in a [Brain tumor classification Kaggle dataset](#) where the slices were already saved as jpg images. I anticipate this will be the most challenging aspect of the project, since it is where my learning curve is greatest.

However, I have already found some online tutorials I am working through. I am eager to learn more about MRI files – including the DICOM and NIFTI formats – and investigate different ways to select and pre-process slices from these files.

Data Leakage Issue

A number of studies have been criticized for having data leakage issues. [13] My understanding of data leakage is that it occurs when images from the training set are also used (or leak) into the test set.

Here's an example: say we have 30 images from each patient and 100 patients, for 3000 images total. If we split the data on those 3000 images, there would probably be some images from a patient in the training dataset and some images from the same patient in the test set. Images from the same patient may be very similar. In effect, we'd be asking the model to "identify" an image it had already seen in the training dataset. That gives the model an unfair advantage.

Instead, I would plan to split the data by patient to avoid this issue.

10 - Final Notes:

I'm very curious about this topic. Increasing my knowledge of and my comfort with working with MRI images would be very helpful to my job. In addition, I'm excited about how this project builds on things I've already done in the GWU data science program. While I'm also a little nervous about forging ahead with my own project, rather than working on one of the preselected ones, I see the following advantages for me to this project:

- Learn a lot about MRI images – how to work with and process them
- Familiarize myself with AD datasets that are commonly used in AD research; this will enable me to be a more critical reader when I read other papers using this data
- I would be able to devote more time to the project if it were a joint school/job project
- Get feedback from two different sources (GWU and my colleagues)
- Possibly learn to use NIH supercomputer biowulf
- Write about a complicated topic in simple language – what I love to do :)
- Very flexible structure – could add other data to the model in the future, such as APOE mutations and/or age and/or CSF and/or other genetic data
- Once I gain familiarity with this simplified 2D modeling, I might want to pursue a similar 3D version of the project in the future.

Thank you for considering this project!

11 - Additional Items from email exchange with Amir

https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images/discussion
https://www.kaggle.com/hyunseokc/detecting-early-alzheimer-s

Lots of interesting stuff to think about! Here are my answers:

https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images

I have a bunch of concerns about this dataset:

- Where did all the images come from? There's no info about how/when it was collected, how the slices were selected, what preprocessing was done, etc.

I plan to work with the original MRI images and clearly document the preprocessing that I do on the NIFTI and/or DICOM files.

- The organization of the dataset is unclear. The 6400 slices were taken from how many patients? Are there multiple slices per patient? Did they take the 6400 slices and then divide them into train and test sets? If, so this could lead to data leakage, i.e. a patient who was included in both the training and test sets.

I would clearly label and document everything – and be sure to divide the train and test sets by patient so that no one person would be in both.

- Looking in the "train folder" under "MildDemented", it looks to me like they only took one slice from each patient, as they are labeled 100, 101, 102.

My project would take multiple slices from each patient.

- On the other hand, looking in the "test folder" under "MildDemented" it looks like they took more than one slice from each patient – here the images are labeled 26 (19), 26 (20), 26 (21), etc. I could be wrong, but since there is no info, we just have to guess.

Again, I would be clear as to where the data was coming from and how it was being used.

- I only see one type of slice used here – from the Axial view (sliced dividing top/bottom).

I would use slices from Axial (top/bottom), Coronal (back/front), and Sagittal (left/right) perspectives. My thought is that would allow us to use more of the 3D aspects of the data, if we looked at each type of image per patient and then combined the results.

- These images appear to have already had skull stripping done, meaning the bone part of the images were removed.

I want to learn more about this. I understand this is necessary for brain segmentation, but is it necessary for a CNN to learn the differences between images? I'm not sure. I might also want to try comparing slices that had the skull stripping preprocessing, as well as using the raw images and compare the models. There are no unprocessed images provided here so no one is able to do that.

- The way the data is presented seems to assume you would create a model that classified slices.

My model would use multiple slices and then ensemble the slices to classify PATIENTS.

https://www.kaggle.com/hyunseokc/detecting-early-alzheimer-s
DETECTING EARLY ALZHEIMER'S | Kaggle
This second dataset doesn't actually contain MRI images, it uses biomarkers generated from the MRI images, such as brain volume. Segmenting the brain and using pre-determined regions of interest (ROI) is a different approach that requires more subject matter expertise. My understanding is that it takes lots of complicated pre-processing to get these ROI measurements. Also, it's not really using computer vision because the researchers are deciding on what the computer will look at in advance, instead of letting the algorithm learn the most important features.

The most similar project I found on Kaggle was:

https://www.kaggle.com/amritpal333/adni4dicomnano10514

This data contains DICOM files – but again, the organization of the data is unclear. If I get the data directly from ADNI, I'm hoping the organization will make more sense.

This project does not contain any actual code notebooks, probably because the files are in DICOM format. I had to take an online class to figure out how to load them, but I did :)

https://arxiv.org/pdf/2002.03419v2.pdf
https://github.com/noxtoby/TADPOLE
https://github.com/ucl-pond/MedICSS-TADPOLE
https://github.com/ssedai026/tadpole-challenge

Interesting! I see a number of differences between the TADPOLE contest and my proposal. From their website:

"TADPOLE datasets include three main types of structural MRI markers of atrophy: 1. ROI volumes 2. ROI cortical thicknesses 3. ROI surface areas, where an ROI (region of interest) is a 3D sub-region of the brain such as the inferior temporal lobe. Obtaining these structural MRI markers from the images is a long and complicated process. This involves registering (i.e. aligning) the MRI images with each other and performing a segmentation of the main brain structures using an atlas-based technique. More information can be found on the Freesurfer website: https://surfer.nmr.mgh.harvard.edu/fswiki/LongitudinalProcessing"

My understanding of this is that they are not using the MRI image files directly – they are using calculated ROI (Region of interest) measurements given in a csv file. Here's what they say about how those measurements are calculated:

"These measures are computed with an image analysis software called Freesurfer using two pipelines: cross-sectional (each subject visit is independent) or longitudinal (uses information from all the visits of a subject). The longitudinal measures are more robust, but the downside is that there are more missing values in our TADPOLE spreadsheet. The MRI biomarkers in TADPOLE can be found in the columns containing UCSFFSX (cross-sectional) and UCSFFSL (longitudinal)."

There's actually a newer version of this software called "FastSurfer."
https://ieeexplore.ieee.org/document/9474779

There's an interesting video about FastSurfer here: https://deep-mi.org/research/fastsurfer/#:~:text=Answer%3A%20FastSurfer%20is%20highly%20reliable,thickness%20measurments%20compared%20to%20FreeSurfer.

But to me the important part is where they say this: "Obtaining these structural MRI markers from the images is a long and complicated process."

In order to avoid this step of feature extraction and make things easier, some researchers have proposed using CNNs instead to classify the images. It has also been hypothesized that dividing a brain into separate "regions of interest" may not be the best way to detect whole brain structural changes.

The 2020 Fang et all paper "Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis" is the most similar one I've come across to what I'm proposing. They also extracted 2D slices from the MRI and PET images, and then used three different pre-trained networks (GoogLeNet, ResNet, and DenseNet) to get a probabilistic score for each image. Finally, they used a Decision Tree to combine the probabilistic scores from each CNN and get their final prediction. The pdf of the paper is attached.

References:

1.   Wyman, B. T. *et al.* Standardization of analysis sets for reporting results from ADNI MRI

      data. *Alzheimers. Dement.* **9**, 332–337 (2013).

2.   Yala, A. *et al.* Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk

Model. *J. Clin. Orthod.* JCO.21.01337 (2021).

3. Alzubaidi, L. *et al.* Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. *Cancers* **13**, (2021).

4. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding Transfer Learning for Medical Imaging. *arXiv [cs.CV]* (2019).

5. Helaly, H. A., Badawy, M. & Haikal, A. Y. Deep Learning Approach for Early Detection of Alzheimer's Disease. *Cognit. Comput.* 1–17 (2021).

6. Li, Y. *et al.* Transfer learning-trained convolutional neural networks identify novel MRI biomarkers of Alzheimer's disease progression. *Alzheimers. Dement.* **13**, e12140 (2021).

7. Tanveer, M. *et al.* Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. *IEEE J Biomed Health Inform* **PP**, (2021).

8. Folego, G., Weiler, M., Casseb, R. F., Pires, R. & Rocha, A. Alzheimer's disease detection through whole-brain 3D-CNN MRI. *Front. Bioeng. Biotechnol.* **8**, 534592 (2020).

9. Fang, X., Liu, Z. & Xu, M. Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis. *IET Image Proc.* **14**, 318–326 (2020).

10. Hon, M. & Khan, N. M. Towards Alzheimer's disease classification through transfer learning. in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 1166–1169 (2017).

11. Prakash, D. *et al.* A Comparative Study of Alzheimer's Disease Classification using Multiple Transfer Learning Models. *Journal of Multimedia Information System* **6**, 209–216 (2019).

12. Tufail, A. B., Ma, Y.-K. & Zhang, Q.-N. Binary Classification of Alzheimer's Disease Using sMRI Imaging Modality and Deep Learning. *J. Digit. Imaging* **33**, 1073–1090

(2020).

13. Yagis, E. *et al.* Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci. Rep.* **11**, 22544 (2021).