Computer Vision Based Automation of Alzheimers Disease Diagnosis Using MRI

Ihsan Baran Sonmez
b21427358@cs.hacettepe.edu.tr

Merve Gul Kantarci
b21527104@cs.hacettepe.edu.tr

Hacettepe University
Computer Engineering Department 06800 Beytepe Ankara

Abstract

Alzheimer's disease (AD) is a neurological disease that affect many people. The early and accurate diagnosis is very important for patients' life. The regular diagnosis process includes brain MRI scans. We aim to automate the diagnosis process with help of the MRI scans by using deep computer vision methods and 2D and 3D data modelling. At the end, the program should be able to classify the MRI scans as cognitively normal or with AD.

1. Introduction

Alzheimer's disease is a progressive neurological disease that its symptoms slowly develop over time. The problems in the transmission of brain signals occurs as a result of the death nerve cells. Alzheimer's disease causes memory and thinking skills loss hence difficulties in daily life. It is highly associated with increasing age, however age is not the only factor. Genetic, lifestyle and environmental factors also effect. In Turkey, approximately 600.000 people are suffering from Alzheimer's Disease. [4] In USA, it is 6th leading cause of death. [3] There is no definite cure for Alzheimer's Disease. However, there are treatments to symptoms in order to increase the life quality. Early diagnosis offers patients to a longer life. This disease is usually diagnosed as a result of psychiatric examinations, but it is quite difficult and not definite, especially at early stages. In addition to psychological examination, Alzheimer's also shows symptoms in MRI results. Having a tool that can make an auto diagnosis helps to accelerate the procedure and raises the confidence in the final diagnosis.

Magnetic resonance imaging (MRI) is a powerful tool to study the structure and function of the brain. We interest in MRI which is a widely-used technique. [1] There are many variants of MRI, and as a result of our research we have observed that the most successful version is T1-weighted MRI. T1-weighted scans provide anatomical images with good contrast between the gray and white matters and it gives more successful results for Alzheimer's Disease. Neuroimaging or brain imaging is a discipline to use of various techniques including MRI to image the structure, function, or pharmacology of the nervous system. [12] With the developing technology, now it has good tools to process the scans, extract information from scans.

In the last years, automating the diagnosis using MRI became more popular. Deep learning methods interferes highly with the medical subjects. Automating the diagnosis of neurological diseases is a convenient and apparently successful application of deep learning. We aim to make a contribution to the field by this study.

2. Related Work

A study made in March 2016 [15] investigates a way to classify Alzheimer's Disease by using MRI with deep learning methods. The dataset employed was provided by Alzheimers Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award numberW81XWH-12-2-0012). They used MRI scans of 28 AD patients and 15 normal people. First 10 slices extracted were removed. In data preprocessing they used: motion correction, skull stripping, Gaussian smoothing. They worked with LeNet-5 [17] with reasons of having a binary classification task but a very complicate one. Data splitted into train (60%) validation (20%) and test (20%) sets with 5-fold cross validation. They trained the network in 30 epochs with the batch size of 64. At the end, they obtained the model with the 96.86% accuracy.

Another study for automated diagnostics of Alzheimer's Disease [9] suggests to use 3D convolutional neural network with 3D MRI data. The motivation was considering the 3D scan as a whole in classification process. They use a dataset of T1-weighted MRI scans of 47 AD patients and

34 Normal Controls from Alzheimers Disease Neuroimaging Initiative (ADNI) [2]. The scans are standartized to 56x56x56 size and only middle cross-section contributes. The data splitted as 66% train data and 33% validation data. The CNN model had three convolutional layers. Each layer had 5x5 filters and followed by max pooling (2x2) layers. They finalized with 0 training loss and 0.02 validation loss. The model is trained with SGD and learning rate is set to 0.001. It stabilized after 40 epochs. It reaches to accuracy of 93% in the test set. The expected future study includes a better data preprocessing and a better CNN model.

Another study [13] also uses dataset from Alzheimers Disease Neuroimaging Initiative (ADNI) [2] with classes clinically normal (CN), mild cognitive impairment (MCI) and Alzheimer's disease (AD). They employ a model with 3 3D convolutional layers and 2 FC layers for the classification. They build a hypothesis that 2D convolution on multi-channel loses the information in the channel dimension. Their results were better compared to previous work that uses 2D convolution in their own metric system. But they say that practical results are not enough also can not be comparable with 100% confidence without and actual cowork to prove the hypothesis and it needs further research is needed to substantiate the hypothesis. The converged model reaches to 81% accuracy on test set, without any misclassifcation between AD and CN classes but misclassification between MCI and AD.

3. Method

The previous works generally employ deep convolutional neural network models. Our approach was similar. We followed two different approaches for 2D data modelling and 3D data modelling so that they can be compared, and compatibility of the task for 3D and 2D modelling can be decided.

3.1. Method for 2D Data

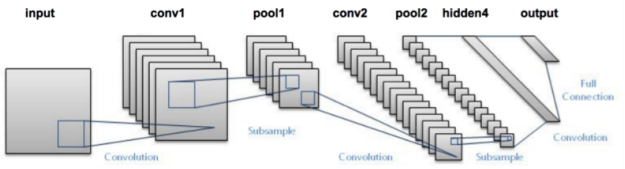


Figure 1: Architecture of LeNet-5

LeNet-5 LeNet-5 was firstly designed by Y. LeCun et Al. in 1998. It sucesfully classifies digits hence it was used in USA to automatically classify hand-written digits for bank checks. Through the it is developed and adopted for other issues. For the modern systems it is a simple network. Previous studies using simple networks for similar tasks got us to decide on it for the first try. It has 7 layers: 3 convolutional layers (C1, C3 and C5), 2 sub-sampling (pooling) layers (S2 and S4), and 1 fully connected layer (F6) and the output layer. But at the end of my studies, I found that this model does not provide enough success in MRI results. So I added 2 new fully connected layers. In addition, I tried to solve the overfit problem by adding the dropout layer. Since Lenet5 was produced to classify 10 different digits and I had only 2 different classes, I updated the shape of the output layer to give 2 results. I set Adam optimizer as the optimizer. I decided it was suitable for this project because the Adam optimizer converged to the desired value in less time. Since this problem is in essence a classify problem, I thought it would be appropriate to use Cross Entropy Loss function in this project. The shape of the MRI results is 3D. Since Lenet5 works on 2D images, I have sliced MRI results on all 3 axes. Instead of training the results of all three axes in a single model, I have trained 3 different models as you can see in the diagram below. The success rate of each model is different and I determined a coefficient according to these success rates for the final calculation stage. As a result of my studies, I have reached the optimum result in this way.

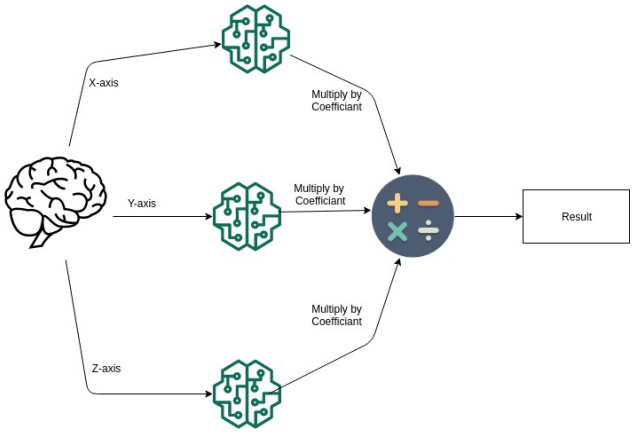


Figure 2: 2D model Architecture

3.2. Method for 3D Data

Early Tries At early stages, first intention was the make dataset convient for 3D U-Net and apply a classification benefitting from this architecture. However, dataset was not suitable for this task, it is needed segmentation cosidering the structure of the U-Net. Also, 3D convolution was very costly in terms of memory. Considering some of the previous works it is decide to start with AlexNet and 3D-ResNet. 3D-ResNet was cancelled due to memory issues, ALexNet did not converge well then it is switched to a shorter custom

Figure 3: 3D Model Architecture

network.

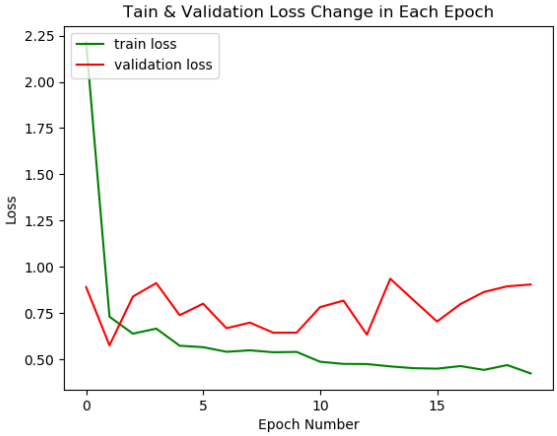


Figure 4: Change of loss during training the model

Final Network First intention was to use 1 channel 3D convolution, however this required to reduce dimensions significantly at the very first layer due to memory issue. Then it is employed another approach to problem: multi-channel 2D convolution. First dimension of input data is given as channel, which being 176 for the scans sized 176x256x256. Final network consists of 3 2D multi-channel convolution layers, each layer is followed by Max Pool layers with $kernel(2,2)$ and $stride = 1$. Convolution layers have 8 filters with $kernel(3,3)$ and $stride = 0$ without padding. It has 2 FC layers at the end mapping to 2 neurons as result: 0 being cognitively normal subject, 1 being subject with AD. Model is trained with SGD optimizer with learning parameter $1e-7$ and cross entropy loss is used. This training resulted train accuracy as 0.78 and validation accuracy as 0.66. Then features part (convolution and max pool layers) is freezed and model's FC layer are finetuned. This resulted 0.90 as best training accuracy and 0.79 as best validation accuracy. It means it memorizes the train data at some point so that we chose the best validation accuracy weights. Overfitting was better after adding a drop out layer however still it overfits at some point.

4. Experiments

4.1. Experimental Setup

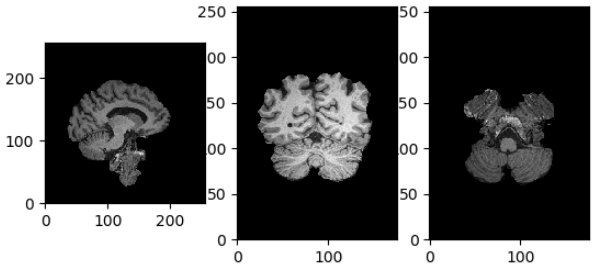


Figure 5: Random central skull-stripped slices from cognitively normal subject

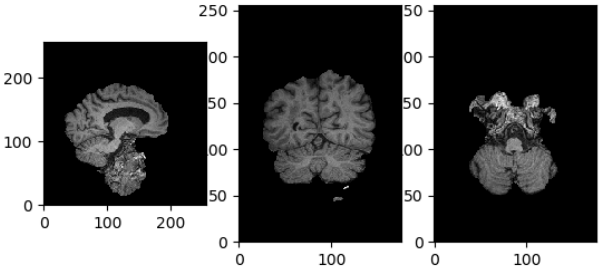
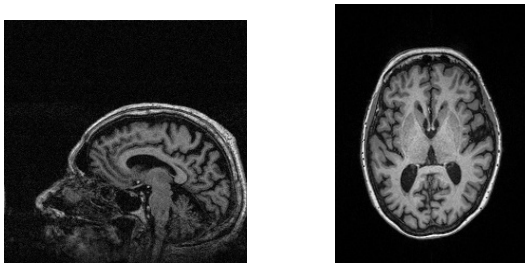


Figure 6: Random central skull-stripped slices from subject with AD

It is prepared a dataset consisting of 178 people with Alzheimer's disease and 178 healthy people. We used the neuroimaging data set of OASIS for these MRI results. The project of Open Access Series of Imaging Studies(OASIS)[11] [5] [14] is making neuroimaging dataset available for the scientific community. They are aiming to accelerate developments in neuroscience in exchange for this totally free service. The MRI results are

in the form of '.nii.gz', a 3D neuroimaging format. It is observed that the models on this subject have been trained with many images in 2D rather than 3D. In order to create 2D images from 3D MRI outputs, we have divided the brain, which is already 3D, into 256 horizontally,256 vertically and 176 depth slices[8]. After these steps, we obtained 686 slices from each MRI (176 + 255 + 255). The number of these slices is not equal for each sample, we observed slight differences in some samples. So we don't know exactly how many slices are used, but according to our calculations we used approximately 352200 slices. But we will not use only 2D samples to train our models, so we have made the 3D versions available for our model. MRI has many different outputs. We decided to use T1-weighted ones. The reason of this decision was that the most common type of MRI's output is T1-weighted in our data set, and T1-weighted contains as much information as the other MRI types.

The results of MRIs are created at many different angles.When we look at the related works on this subject, we have seen that there is no limitation. For this reason, we decided to use all the angles in our project. You can see the example of different angles from Figure 3.

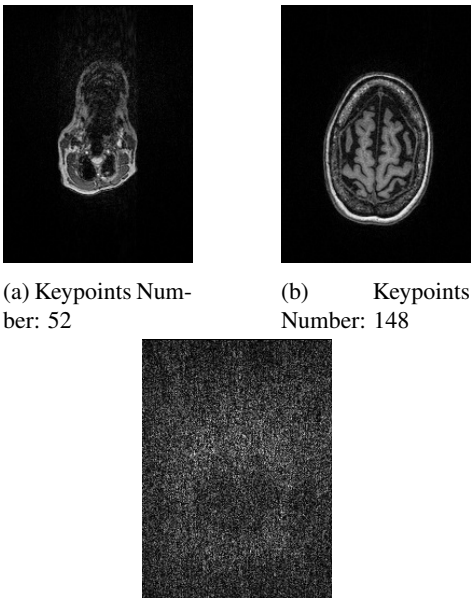


(a) Vertical Slice (b) Horizontal Slice

Figure 7: Example of different angles

Since MRI is standard in size, but the size of the brains is not standard, we observed that there are many empty, noisy slices in the MRI results. You can see an example of this in Figure 4. We used SIFT to overcome this problem. If the number of keypoints from a picture is less than 60, it is likely that it does not give valuable information so that we decided not to use this picture for our model. We also check the slices next to the slice to avoid possible confusion. In this way, we eliminate the effect of noisy data on our models.

There are many different devices that produce the MRI result. We decided to use only the results from the 3.0t device to eliminate the scanning differences between these devices. When we look at related works on this issue, we have seen that the focus is on the results of a particular device.In this project we will use 60% of all data for train, 20% for validation and 20% for test.



(a) Keypoints Number: 52 (b) Keypoints Number: 148

(c) Keypoints Number: 21

Figure 8: Difference of Keypoints number change in varios MRI slice

3D spesific preprocessing It is harder to manage 3D data since it is hard to visualize, one can not make operations with confident compared to 2D. For MRI case it is extremely important that brain regions allign for each image. Every instanse of the data consists of same specialities (3.0T Scanner, T1-Weighted MRI) that is selected at the begining. When the results narrowed to Alzheimers Disease subjects without anyother known diagnosis, and Cognitively Normal subjects without any known diagnosis, it results a very large set for Cognitively Normal and 178 subjects for Alzheimers Disease case. Some subjects have more than one run for MRI. Any of the MRI data with size of 176x256x256 is accepted to next preprocessing step. Other sizes are rejected because we could not know the source of the difference for sure hence no recovery was possible. Some scans were sized as 176x240x256, the soruce of this reduction was unknown and it had fewer examples compared to 176x256x256 sized data. Also some of the data was 175x256x256 sized, which means we have to reduce one the channel size for other rest and the other part had more dominant distribution. It is prefered to keep the channel size as 176, while throwing 1 or 2 subject with 175x256x256 sized MRI scan.

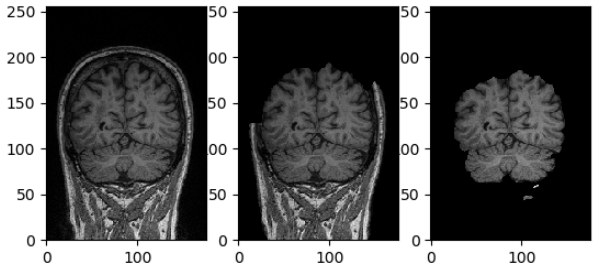


Figure 9: Skull Stripping with raw, frac = 0.2, frac = 0.5

Next step was making the skull stripping. Skull stripping is a critical step. A raw scan includes the nose, neck and skull regions as well. This had to be reduced to get the brain region only. There are neuroimaging tools for extracting the skull region and brain region. We only needed the brain region. Brain Extraction Tool (BET) [16][10] is a very useful tool for this job. It does not provide a very user-friendly GUI but it is very simple, straightforward and also easy to manage from terminal. It does not support Windows so that I had to extend my working environment to Ubuntu as well. It can be automated for a series of NiFTI files using nipype [7] library in Python. It was a huge plus for us and it made the process significantly faster, writing mannual command would be very time-consuming. Fractional intensity threshold (between 0 and 1) is an important setting which decides to width of brain outline. Smaller values might result not a good strip, as it is seen from the figure. Default setting as $frac = 0.5$ was suitable for our case. After skull stripping was done with all MRI scans, another run to make sure of the sizing was necessary. However as expected, skull stripping did not change the size. From the final skull stripped set, 120 subjects with AD and 120 cognitively normal subjects are chosen randomly for train set. 24 subjects for validation set and 24 subjects for test set are chosen. Cognitively normal subjects were more than this total but it is aimed to make a uniform distribution over the classes.

Environment We used Python 3.7 for the programming part. We used GitHub as a version control system and preprocessing part of the 3D implementation and 2D modelling are developed on Ubuntu platform, 3D modelling is developed on Windows 10. GPU support was not available due to memory issues. The GPUs we have for this task are Nvidia gt940mx 2gb and Nvidia gtx1050 4gb. Google Colab[6] was harder to use since the data is massive and uploading through Google Drive was very time-consuming.

4.2. Experimental Results

4.3. 2D Model

	Precision	Recall	f1-Score	Samples
Healthy	0.86	0.79	0.83	24
AD dementia	0.80	0.87	0.83	23
Accuracy			0.83	47

Table 1: Precision, Recall, f1-Scores and Accuracy Scores

In the 2D model, as I mentioned before, I used a separate model for each axis. This resulted in unreliable accuracy during training. I used the test data to test the success of the model. The test set consisted of 23 healthy people and 24 Alzheimer's patients. The hyper parameters that I use and get the best results in the model, optimizer = Adam, Loss Function = 0.00014, and the epoch number = 20. The figure below shows the confusion matrix results.

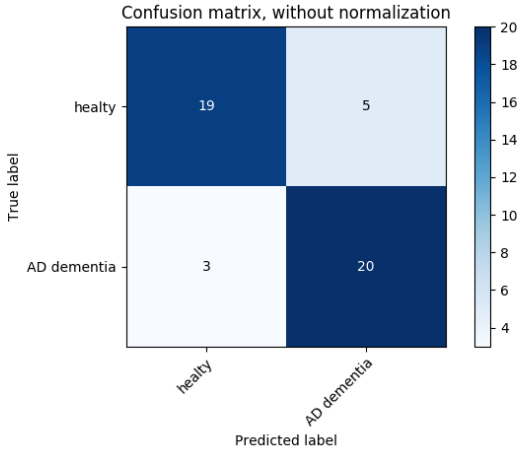


Figure 10: Confusion Matrix of 2D model, without normalization

As you can see in the table above, the success of the model is very close to each other. In addition to this matrix, precision, recall values can be found in the table below. For the best case, we want precision and recall to be equal to 1. Although the results of our model are not extremely good, the high precision and recall values for both cases indicate that the model has learned successfully.

4.4. 3D Model

	Precision	Recall	f1-Score	Samples
Healthy	0.71	0.83	0.77	12
AD dementia	0.80	0.67	0.73	12
Accuracy			0.75	24

Table 2: Precision, Recall, f1-Scores and Accuracy Scores

Addition the validation set, test set is used to measure the success of trained network. Test consists of 12 subjects with AD and 12 cognitively normal subjects. The best model is obtained with SGD optimizer, with cross entropy loss both in first training and fine-tuning as it is mentioned in Method section.

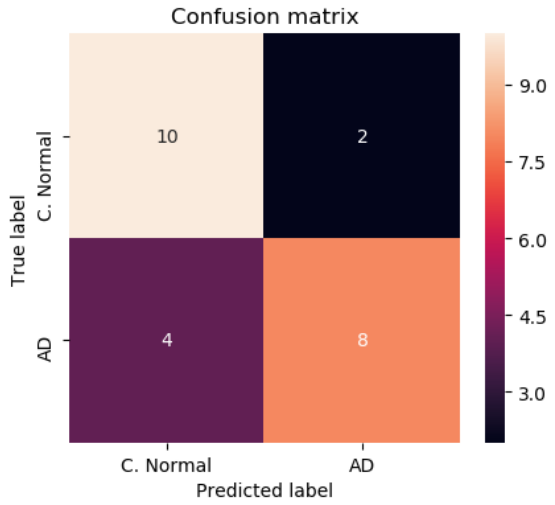


Figure 11: Confusion Matrix of 3D model, without normalization

Model's precision, recall and f1 score values are shown on the table 2. As it is seen from the heat map, model tends to classify the subjects as cognitively normal compared to its AD classification rate. It is more rare that it classifies a cognitively normal subject as AD patient. The results are worse than 2D model's results as it can be reviewed.

5. Conclusion

Neuroimaging has its own properties and own methodology. Its tasks differ from regular image classification task. Neurobiology background is essential for more complex problems. Neuroimages are more difficult to work with than normal images. In order to work on neuroimages, it has to go through some preprocesses. However, Neuroimages do not consist only of the sections we want. Therefore, we need to eliminate these regions for our model to be more successful. To overcome this problem, we also experienced SIFT and Mask implementation. We also used a lot of libraries and networks. In particular, we experienced the PyTorch library to build and train the model and the scikit.metrics library, which provides the statistical output we use for the results.Tools for these tasks are not supported on all platforms, sometimes heavy in terms of memory and requires topic-specific knowledge at some point. These being said, feature extraction is improving and has already very satisfying techniques.

Both models result are very promising and 2D modelling gave better accuracy value in this study. There is a difference of 0.08 between two models. This can be considered as a good improvement, but for a certain claim on it, it has to be ensured by a strong hypothesis. Code that is developed for this task shared in GitHub repository. Information of required libraries can be found at the repository. **Data is not shared due to policy of OASIS[11].**

Comparison between models 2D modelling is more common as far as we observed, but the recent lean is on 3D convolution with developing computation power and memory solutions. In theory, 3D modelling uses the data as a whole chunk, so that it seems more suitable. 2D modelling uses slices hence it needs slicing first and needs to combine different networks. 2D modelling makes more weight updates during training. In practice in this study 2D modelling was better. 2D images are easier to manipulate, visualize and one can reject the non-informative parts in an easier way. 3D data might require unique libraries at some point and it requires more memory. Since 2D modelling combines three different networks, code can run on parallel.

Future Work MRI data is very sparse. Sparsity effects badly the performance of the model. Solving the sparsity issue is a good challenge for this task. For Alzheimer's Disease it is not common to make a segmentation but there are previous works that have done it. By preparing a convenient dataset, segmentation can be an option. 3D modelling can be improved by using autoencoders. 1-Channel 3D convolution should be tried. This is possible with using a cloud system or using correct and size-reducing filters at the very beginning. Even with the 3D data, redundant regions can be extracted or effect of some regions can be increased. However, this should be a generalizing condition so that it still remains automated as a computer vision task. Automation is the essential goal here.

References

[1] Neuroimaging as a case study in research data management, February 2018. 1

[2] Alzheimers Disease Neuroimaging Initiative (ADNI). *Available at http://adni.loni.usc.edu/ .2019-05-19*. 2

[3] Alzheimer's Association. Facts and figures. *Avaliable at https://www.alz.org/alzheimers-dementia/facts-figures .2019-05-18*. 1

[4] Turkiye Alzheimer Dernegi. Haberler. *Avaliable at http://www.alzheimerdernegi.org.tr/haber/turkiyede-600-000-aile-alzheimer-hastaligi-ile-mucadele-ediyor/ .2019-05-18*. 1

[5] A. F. Fotenos, A. Z. Snyder, L. E. Girton, J. C. Morris, and R. L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and ad. *Neurology*, 64(6):1032–1039, 2005. 3

[6] Google. Google colaboratory. *Avaliable at https://colab.research.google.com/ .2019-05-18*. 5

[7] Krzysztof Gorgolewski, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*, 5, 08 2011. 5

[8] Harvard. Freesurfer. *https://surfer.nmr.mgh.harvard.edu/ .2019-05-18*. 4

[9] Li X.C. Luo, S.H. and J.M. Li. Automatic alzheimers disease recognition from mri data using deep learning method. *Journal of Applied Mathematics and Physics*, 5(11):1892–1898, 2017. 1

[10] M. Pechaud M. Jenkinson and S. Smith. Bet2: Mr-based estimation of brain, skull and scalp surfaces. *In Eleventh Annual Meeting of the Organization for Human Brain Mapping*, 2005. 5

[11] Oasis-brains. The open access series of imaging studies. *Avaliable at https://www.oasis-brains.org/ .2019-05-18*. 3, 6

[12] Wikipedia [online]. Neuroimaging. *Avaliable at https://en.wikipedia.org/wiki/Neuroimaging . 2019-06-22*. 1

[13] Bouke Regnerus and Raoul Fasel. Classifying alzheimers mri scans using 2d and 3d convolutional neural networks. *Avaliable at https://github.com/regnerus/keras-alzheimers-3d-conv/files/2111208/main.pdf . 2019-06-22*. 2

[14] Miller JP Kinscherf DA Grant EA Morris JC Berg L. Rubin EH1, Storandt M. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Arch Neurol*, 55(3):395–401, March 1998. 3

[15] Saman Sarraf and Ghassem Tofighi. Classification of alzheimers disease using fmri data and deeplearning convolutional neural networks. pages 0–5, 2016. 1

[16] S.M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, November 2002. 5

[17] Y. Bengio Y. LeCun, L. Bottou and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):22782324, 1998. 1