Deep Learning Seminar Project: Three-Way Alzheimer's Disease Classification

Aly Valliani AVALLIA1@SWARTHMORE.EDU
Swarthmore College, 500 College Avenue, Swarthmore, PA 19081 USA

Andrew Gilchrist-Scott AGILCHR1@SWARTHMORE.EDU
Swarthmore College, 500 College Avenue, Swarthmore, PA 19081 USA

Abstract

Alzheimer's Disease (AD) is neurological disease characterized by a loss of memory and other important mental functions. Current techniques for AD diagnosis are inaccurate and there exist no cures upon disease onset. Therefore, a systematic method of disease diagnosis is of immense importance.

This paper describes the deep learning techniques we utilized in order to improve upon the state of the art accuracy for 3-way AD diagnosis. We examined the work and methodology of several notable authors in the field of computer vision and sought to apply their principles to medical images, namely structural brain MRIs, provided by the AD Neuroimaging Initiative. Results indicate that our architectures were unable to match the current state of the art (Payan & Montana, 2015; Suk & Shen, 2013). We attribute the presented findings to our model's inability to extract proper features for classification due to the high dimensionality of the data.

1. Introduction

Alzheimer's Disease (AD) is a neuro-degenerative disease and the sixth leading cause of death in the United States, affecting 5.3 million people (Alzheimer's Association, 2015). It is characterized by a progressive loss of memory and other important cognitive functions (Reiman et al., 2011). To a first approximation, the disease is caused by the degeneration and death of brain cells in the hippocampus, which initiates atrophy of the entire brain. However, the specific biological mechanisms underlying the onset of this disease remain unclear (Shaffer et al., 2013). Research has, however, implicated abnormal neuritic amyloid beta (Aβ) plaques and neurofibrillary fibers that function to hinder neuronal communication and disrupt cellular processes as the potential biological bases of this disease (Goedert et al., 2006).

While there is currently no cure, a large amount of research is being devoted to preventative and palliative treatments (Ramirez et al., 2005; Rolland et al., 2008; Langbaum et al., 2014). However, diagnoses for early detection are inaccurate, which significantly impedes research on preventative treatments (Reiman et al., 2011). Therefore, a robust automatic diagnosis of patients, especially one which could successfully predict the onset of full AD, would be extremely clinically useful. Several systems are currently under development to solve this problem, but near perfect diagnosis (greater than 99%) is necessary for any system to gain clinical acceptance.

Most automated techniques for AD diagnosis employ computer vision techniques to examine structural magnetic resonance images (MRIs), which allow the inspection of the physical layout of the brain. A novel examination technique (which we'll expand upon in Section 2) is to use convolutional neural networks (CNNs) and other deep learning techniques in order to examine hierarchical features of the images (Gupta et al., 2013). However, very few of these papers have combined multiple CNNs into any depth, which other work (Krivhesky et al., 2012; LeCun et al., 2010; Szegedy et al., 2014) indicates might be extremely useful for image classification. In addition, a system optimized for classification of AD would likely have good transfer learning potential to other neurological disorders, including but not limited to schizophrenia, white matter lesions, and other forms of dementia.

Currently, the majority of work has focused on binary classification for patients with AD versus those who are cognitively normal (CN) (Kloppel et al., 2008). However, this

CPSC 93 Deep Learning Reading Group Final Report Proceedings, Swarthmore College, Fall 2015.

represents a gross simplification of the clinical diagnosis of AD. The disease exists along a spectrum with one side being cognitively normal and the other being full, debilitating AD which can currently only be truly confirmed at autopsy. A better approximation of this spectrum is 3-way classification between AD, CN, and a third category that signifies some mental impairment prior to the full onset of AD symptoms: mild cognitive impairment (MCI). This disease state can further be split into early and late versions (EMCI and LMCI, respectively), but the three classes are usually deemed to be diagnostically sufficient (Zhang et al., 2011). Therefore, the task can be formulated as follows: **Given a corpus of MRI images labeled as CN, MCI or AD, build a model that accurately conducts 3-way classification (CN vs. MCI vs. AD) for novel MRIs.**

In this paper, we describe our approach to the above-mentioned task. More specifically, we design a deep architecture that aims to match or best current state-of-the-art 3-way AD classifiers. This paper is organized as follows. Section 2 examines existing work in the field and the theoretical underpinnings of popular deep architectures. Section 3 describes the utilized data set. Sections 4 and 5 present our architectures and the preprocessing pipeline that was used for 3-way classification. Sections 6 and 7 provide and discuss the results obtained from the aforementioned models. Sections 8 and 9 conclude with avenues for future work and a meta-discussion of our project.

2. Related Work and Theory

Until 5 years ago, most moderately successful AD classification came from custom engineered features fed into standard classifiers such as support vector machines (SVMs). These methods resulted in decent binary classification of AD vs CN that surpassed the accuracy of specialized radiologists (Kloppel et al., 2008). However, the methods were unsuccessful in 3-way classification (Zhang et al., 2011). Most of these findings relied on the discovery of regions of interest (ROIs) which were thought to be indicative of AD. However, the definition of these ROIs were either predicated on a pre-existing atlas of the brain (Zhang et al., 2011) or on a model which assumed some set topology of the brain (Ortiz et al., 2014). Both of these can become invalid assumptions in the presence of extreme AD, since the disease causes significant deformation of the brain away from the healthy norm (Gupta et al., 2013).

Recent work has focused on automatic feature extraction through pattern recognition that expands throughout the brain (Gupta et al., 2013). This is accomplished primarily through auto-encoders (AEs), convolutional neural networks (CNNs) and restricted boltzmann machines (RBMs).

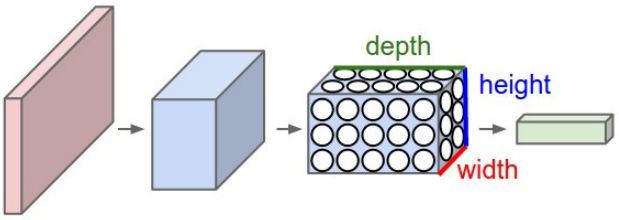

Figure 1. Diagram of a simple convolution layer. The CNN transforms the input image (in RGB) into a vector of resulting values from the application of filters and an activation function. It can also be noted that this vector has a depth corresponding to the number of filters which were placed on the image slice. This image was provided courtesy of (Karpathy, 2015).

2.1. AEs

Auto-encoders are a peculiar variety of artificial neural networks in which both the input and output are the same. Through this architecture, the model attempts to learn a compact representation of the data in a manner similar to principal component analysis (PCA). However, when random noise is added to the input but not the output, the model is forced to learn a high order representation, enabling us to use an over-complete encoding (the size of the encoding is larger than that of the input) that can encompass some of the larger patterns present in the distribution of all inputs (Vincent et al., 2008). This process is called denoising. The use of an AE to initialize each level of a deep neural network can significantly improve overall accuracy, since the AE can behave as a generalizer to the true distribution of the possible inputs (Erhan et al., 2010). Throughout our discussion, all AEs are assumed to be denoising.

The first large, successful use of a deep network with AEs achieved state of the art accuracy for both 3-way classification and MCI to AD conversion, a related sub-problem (Suk & Shen, 2013). The work used AEs as a feature selector to feed vectors into a multi-kernel SVM. Later work (Gupta et al., 2013) used AEs as an initializer of weights for an artificial neural network (ANN).

2.2. CNNs

Developed in the mid to late 1990's, CNNs are a method of iteratively and automatically selecting filters of interest that can be useful to classify images (LeCun et al., 1998). The idea of using convolutions (also called filters or kernels) is built upon the concept that individual patterns found in certain areas of the image can give insight into content. While many successful works have been done that use very simple, hand-designed convolutions, complicated image recognition requires a large breadth of filters, as can be seen in Figure 2. These filters are learned in the method

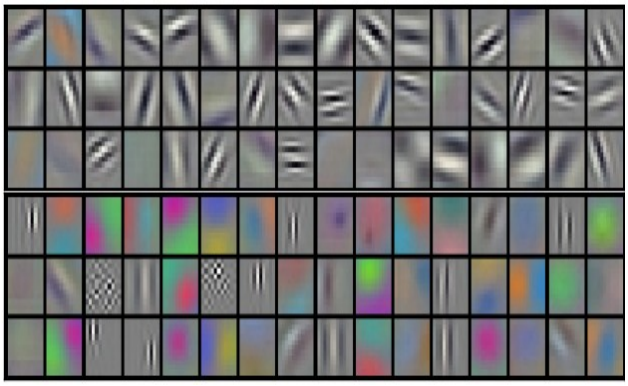

Figure 2. Example filters that were learned by (Krivhesky et al., 2012). Each of these filters is 11x11x3, i.e. they can stretch over an 11x11 pixel region in three colors

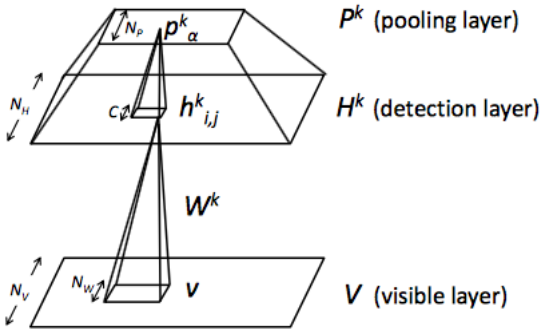

Figure 3. Diagram of a convolutional RBM. Hidden units contained within the hidden detection layer interact with a local region of visible units. Importantly, the weight matrices mediating this interaction for a local region of hidden units are the same, which is reminiscent of CNNs. This image was provided courtesy of (Lee et al., 2009).

of a neural network, whereby a series of weights (the filter) are convolved over an image to extract information in a series of layers. This information is fed into a neural network classifier, which produces a probability distribution over the label set. All errors in this process are propagated back through the network to adjust the weights in the next iteration. While each section of the image has filters that are unique to it, we generally tie weights together such that we have a single filter that is moved across the entire image. Mathematically, this is roughly equivalent to saying

$$a = f\left(\sum_i w_i x + b\right) \tag{1}$$

where a is the activation (the result of the filtering), f is some activation function, w_i is the filter, b is some bias term, and x is some chunk of the image. The a values form a resulting convolved image. The filtering layers, or convolution layers, can be stacked on top of each other. The system's general form can be visualized in Figure 1.

Later work by (Gupta et al., 2013) took the idea of CNNs and AEs and applied it to brain images. The researchers first gathered a patchwork of small 8x8 sections of MRIs and natural images (i.e. standard 2D images that are not necessarily medical in nature) and employed an AE to initialize filters on a CNN. This CNN was then trained in a single layer on patient MRIs to first conduct binary classification between all three classes followed by three way classification. They achieved state of the art accuracy, which was only eclipsed more recently by another CNN, which used full 3D filters (Payan & Montana, 2015).

2.3. RBMs

Restricted Boltzmann machines are a two-layer, bipartite, undirected graphical model that aims to learn a probability distribution over its set of inputs. At a high-level level, it can be considered similar to a generative artificial neural network with visible units fully connected to hidden units. Functionally, it's similar to an autoencoder in that it can be used to a pre-train a network, such as a CNN, in an unsupervised manner. The RBM can be considered as a building block to deep belief networks, which can be conceptualized as stacked RBMs. Such models have been analyzed for neuroimaging purposes and generated results exceeding methods routinely used in the field (Plis et al., 2014)

A convolutional RBM (CRBM) is a hybrid between a CNN and RBM and can intuited as an ANN with shared weights between visible and hidden units as depicted in Figure 3. This property enables the CRBM to function as an edge detector for image-based data (Lee et al., 2009). Although to our knowledge such networks have not been analyzed for neuroimaging studies, they present an interesting alternative that require further research.

3. Dataset

All utilized functional MRI data was obtained as part of the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a longitudinal, multi-site study aimed at determining the biomarkers associated with AD onset (`adni.loni.ucla.edu`). The study began in 2004 with 200 CN, 400 MCI and 200 AD subjects.

For the purposes of this study, two datasets were utilized. The first was a small subset of images that consisted of 102 CN, 203 MCI and 92 AD images, and represented the final MRI test for all patients who remained in the study throughout its three year duration. The second, which we obtained later in our research, was a larger dataset that con-

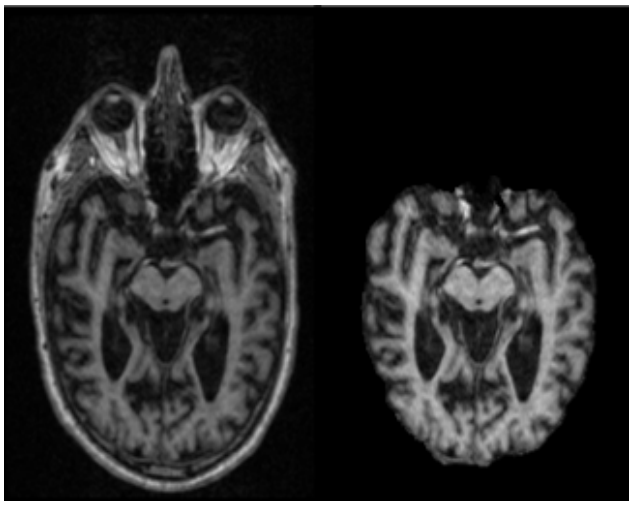

Figure 4. A transmedial slice of a brain before (left) and after (right) skull stripping

sisted of 220 CN, 294 MCI and 197 AD images, which represented the initial screenings for all candidates in the study.

4. Architecture

4.1. Preprocessing

We applied the following three standard pre-preocessing steps to the MRIs to enable ease of use within our pipeline: normalization, coregistration, and skullstripping. Normalizing was to the range of $[0, 1]$.

4.1.1. Coregistration

Coregistration was utilized to rotate and scale the brain images so that they all existed in roughly the same space. This technique corrects for the fact that different MRI machines record the images as different sizes and with different orientations. While convolutions as a mathematical procedure are translation invariant, they are not able to capture all scaling and rotational disparities between images. We performed all coregistration using statistical parametric mapping (SPM), a commonly used tool for brain research (Penny et al., 2006). The template image to which all other images were registered was chosen arbitrarily from the training set.

4.1.2. Skull Stripping

Skull stripping is the process of removing all non-brain material from the MRI. Theoretically and experimentally (Gupta et al., 2013), this tissue has no positive affect on the discovery of the underlying diagnosis but may by chance appear to be similar to a significant feature from the brain's interior. The skull stripping was performed using the FreeSurfer software package (Fischl et al., 2002) and example results can be seen in Figure 4.

4.2. Data Modification

After preprocessing, we performed a series of different operations on the data to make computation possible on standard amounts of memory. A single MRI, when uncompressed, fills approximately $\frac{3}{4}$ GB of memory, so it was necessary to design different computational strategies to account for the inability to store our entire dataset simultaneously on RAM.

4.2.1. Transmedial Slices

Many of our tests involved using only a single, 2D slice of the brain, which worked well with many pre-existing deep learning frameworks, namely Pylearn2 (Goodfellow et al., 2013). These slices are transmedial cuts at both ears and eyes and incorporate many of the significant symptoms of AD including enlarged ventricles and reductions in cortical mass.

4.2.2. Layered versus flat

We took two different approaches to represent the multi-dimensional image. One was to treat the brain as a multi-channel 2D image, equivalent to saying each of the layers in the Z-dimension were like colors in a standard RGB image (Ypsilantis et al., 2015). The other option was to lay the Z-dimension layers of the brain flat along the X-dimension, such that the data appeared as a very long 2D image.

4.2.3. Downsampling

Regardless of the choice in Section 4.2.2, the dataset was too large to fit entirely on RAM. We were forced to either take a 90 patient subset of our data or downsample to approximately one fifth of the original size. Downsampling was achieved through standard computer vision downsampling techniques available through OpenCV (Bradski).

4.2.4. Streaming

To combat the size restrictions imposed by RAM, we designed a streaming protocol that trained on a single image at a time then returned it to disk. Though this is theoretically better than the previous data modification options, it took enormously longer to train and allowed for less model optimization.

4.3. Use of AEs

For those architectures that employed an AE to initialize the first layer of the CNN, we chose to take a similar ap-

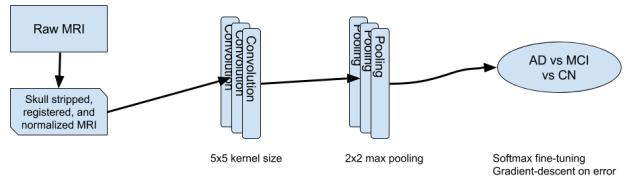

Figure 5. The one-layer CNN model. The convolution layer had 128 filters.

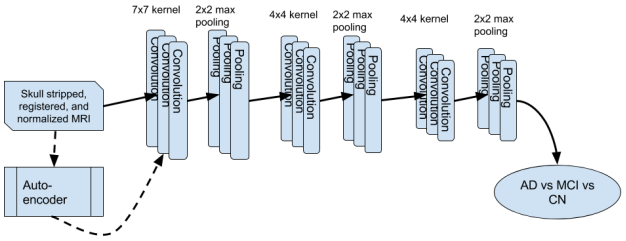

Figure 6. The three-layer CNN model. Both the large and small three layer models had the same topology.

proach to Gupta et al. 2013. We began by randomly selecting 100,000 8x8 patches from the dataset. These were used to train a single-layer convolutional auto-encoder. The weights of this AE model were then used to initialize the first layer of a large CNN.

4.4. CNN Structure

We tested three size CNN's: a one layer model (Figure 5), a three layer model (Figure 6), and a large three layer model. The latter two were also tested with the first layer initialized by the patch results of the convolutional auto-encoder. The large three layer model had 64, 128, and 64 filters at each level, respectively, whereas the smaller model had 6, 12, and 6 filters at each level.

4.5. RBM Structure

Since the Pylearn2 library did not implement a functional CRBM, we constructed our own. Prior to running the constructed architecture on the MRI data, we validated its proper functionality on the classic MNIST data for handwritten digit recognition. A misclassification error rate of $<10\%$ was obtained, which corroborated the architecture's functionality.

The final architecture was a single layer model with 11x11 convolutional filters trained with 2x2 probabilistic max pooling. The algorithm minimized the contrastive divergence (CD) cost function as is standard procedure for RBMs and its variants.

Model \ Dataset	Small	Large
1-Layer CNN	.48	.50
3-Layer CNN	.48	.74
3-Layer CNN with AE	.70	I
CRBM	.48	.50
SVM	.75	I
SVM with PCA	X	X
2-Layer ANN	.66	.75

Table 1. All notable model errors. Note that these are not under consistent hyper-parameters such as learning rate or noise percentage, but rather the highest achieving model in each category. Results involving PCA were incalculable on the machines available and therefore are denoted with an 'X'. All models that have not yet completed training are denoted with an 'I'.

Class	Misclassification Error
AD	72.41%
MCI	71.74 %
CN	68.00 %

Table 2. Individual class misclassifications for a run of the 3-Layer CNN with AE that achieved a 71% error overall.

5. Baseline Architectures

5.1. SVMs

Slightly following the successful model of Suk & Shen 2013, we fed our preprocessed images into a SVM, both with and without dimensionality reduction by PCA. This provided a naive baseline for how well our models should be performing.

5.2. Deep ANNs

CNNs are appealing for image classification because they respond to an image's internal structure and acknowledge that a pixel value at one coordinate influences the distribution of the values of its neighbors. In contrast, a normal ANN makes no such assumption, and any consistent reordering of the input should produce statistically identical results.

6. Results

Since we attempted many models and variations on the data, we report here only our best results (Table 1). In particular, we tracked the learning rate of the 3-layer CNN, its batch train error, and the reconstruction error associated with the auto-encoder, all of which can be seen in Figure 7, Figure 8, and Figure 9, respectively.

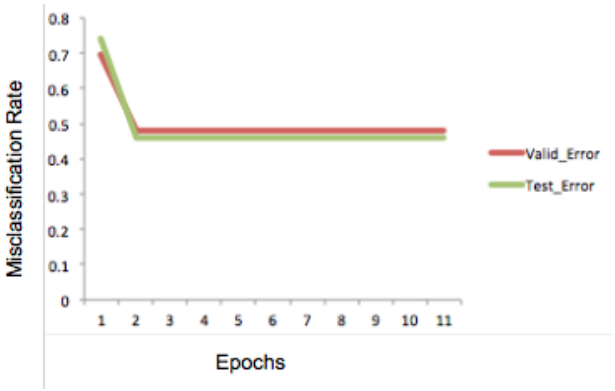

Figure 7. Learning curve of the 3-layer CNN over 11 epochs on the small dataset. Both training and testing error plateau at 2 epochs, thereby indicating that the model is quickly converging to a sub-optimal local minimum.

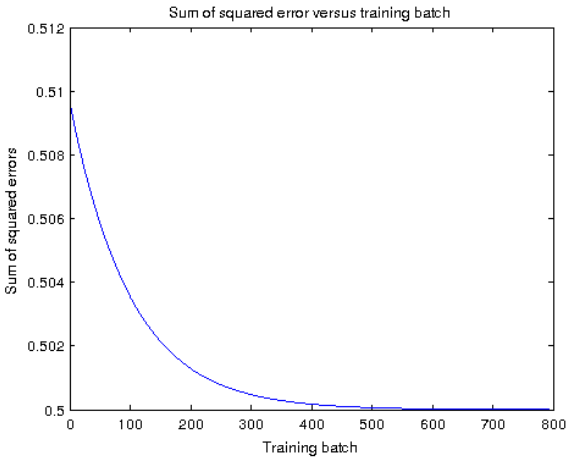

Figure 8. Training error curve of the 3-layer CNN with AE over 8 epochs. Each batch was 3 image large, test performed on the smaller dataset. It should be noted that the y-scale is very small on this image, indicating very little change in error over the course of training.

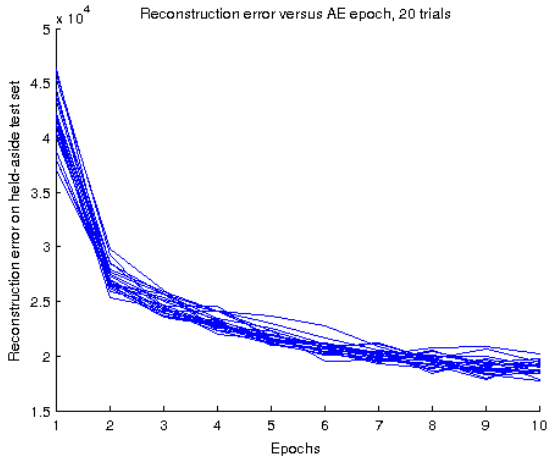

Figure 9. Reconstruction error for the AE built on 100,000 8x8 2D patches of the ADNI dataset. As these 20 trials showed, most error plateaued after the 9th epoch, thus 10 epochs of AE training was used for all later models.

7. Discussion

Table 1 clearly shows that our results were very poor in virtually every category. This was with the implementation of nearly every feature outlined in the previous sections, including though not limited to:

- Dropout
- Layered versus flat images (Section 4.2.2)
- Downsampled versus full-size images (Section 4.2.3)
- Early stopping
- Taking subsections of the data model
- Balancing classes in training and testing sets
- Maxout activation function

None of these methods had a statistically significant effect on the results. In fact, the large swing of results in Table 1 are primarily from the underlying distributions of the randomly separated test set. Results such as the 1-layer CNN, the SVM, and small set 2-layer ANN all stem from labeling all images as the most common class (usually MCI). Those models that did seem to "learn" anything such as the 3-layer CNN with AE were unable to find any meaningful patterns, and nearly immediately found and remained in a local minimum. This is supported by Figure 7 and Figure 8, which show that the model hardly trains beyond the first few epochs. Additionally based on Table 2, there are no obvious patterns to preference one class over another (aside from the fact that the architecture does worse than chance).

These results as a whole lead us to believe that our data is too high dimensional to successfully be approached by our methods. This has some support in the fact that the widely successfully models in 2D image recognition and segmentation (Szegedy et al., 2014; Krivhesky et al., 2012; Zheng et al., 2015) are significantly deeper than ours (> 10 convolutional layers) with many more filters (on the order of 2^{10}). The curse of dimensionality would explain our results of either settling very quickly into a bad local minimum or making no sense of the data. While we would hope that any amount of automated feature selection (essentially the purpose of deep learning) would have reduced this problem, it is understandable that with a dataset that has at most 711 examples and 10,878,976 dimensions (256x256x166 MRIs) we may see some problems. It should be noted that this dimensionality was so large that common reduction techniques such as PCA were impossible with any significant number of images (> 30 of each class) even when using singular value decomposition (SVD), the most memory efficient computation strategy.

8. Conclusion

This paper discussed the various deep learning architectures we utilized for Alzheimer's disease classification. Using the foundational work of other researchers in the field of neuro-imaging, we aimed to improve upon existing classification accuracies by building deeper architectures, utilizing methods for unsupervised pre-training and optimizing model parameters. However, our architectures fell short of existing techniques with classification accuracies in some cases comparable to shallow architectures. These results indicate that the high dimensionality of the data prevented adequate feature extraction and learning.

In the future, we will focus on using better supported libraries (specifically PyCaffe (Jia et al., 2014)) to implement larger and deeper networks with the use of a GPU to make computation tenable. Additionally, future work will likely begin with the exploration of related classification tasks, such as the simpler task of 2-way classification (CN vs. MCI, CN vs. AD, MCI vs. AD). It is our hope that the simpler task will provide insight into the inadequacies of our various architectures. For example, poor performance on 2-way classification between CN and AD brains would indicate fundamental issues with our architectures, since such a task should be relatively easy to perform. In contrast, poor differentiation of MCI and AD brains would indicate issues with feature extraction and would warrant the use of increasingly deeper networks, enhancements in unsupervised pre-training using natural images (Gupta et al., 2013) and the incorporation of positron emission tomography and cerebrospinal fluid data.

9. Project Remarks

We encountered various challenges throughout the course of this project. The biggest hurdle involved dealing with Pylearn2's lack of adequate documentation and functional limitations. Although the library's stock models perform well on classic datasets, such as the MNIST and CIFAR10 datasets, they don't provide the necessary flexibility required when working with novel data. As such, the library is great for users looking to tinker with deep learning, but is not optimal for those interested in pursuing independent research. It is unsurprising that the library has recently been abandoned by all of its previously active developers. MATLAB's Deep Learning Toolbox was able to fill in the functional gaps, but computation was painfully slow due to the library's lack of GPU support.

Our issues with existing libraries is indicative of Deep Learning's relatively recent resurgence. Many libraries seem to lack the necessary functionality required for proper research. We hope to continue working on this project with the proper research libraries in the hopes of improving Alzheimer's disease classification and remaining at the forefront of this surging field.

10. Acknowledgements

We would like to acknowledge the infinite help of Ameet Soni, our primary adviser for this project, in addition to Lisa Meeden and Rich Wicentowski for organizing a course that has introduced us to an emerging and powerful field in machine learning. We also thank Professor Jeff Knerr for his technical assistance with setting up the various deep learning libraries for our use. And last but certainly not least, we thank our fellow reading groupers, Klarissa Khor, Ravenna Thielstrom and Razi Shaban, for their humor, feedback and suggestions.

References

Alzheimer's Association. 2015 Alzheimer's disease facts and figures. *Alzheimer's Dement*, 11(3):332–384, 2015.

Bradski, G. *Dr. Dobb's Journal of Software Tools*.

Erhan, Dumitru, Bengio, Yoshua, Courville, Aaron, Manzagol, Pierre-Antoine, Vincent, Pascal, and Bengio, Samy. Why Does Unsupervised Pre-training Help Deep Learning? *Journal of Machine Learning Research*, 11: 625–660, 2010.

Fischl, Bruce, Salat, David, Busa, Evelina, Albert, Marily, Dieterich, Megan, Haselgrove, Christian, van der Kouwe, Andre, Killiany, Ron, Kennedy, David, Klaveness, Shuna, Montillo, Albert, Makris, Nikos, Rosen, Bruce, and Dale, Anders. Whole brain segmentation:

Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33:341–355, 2002.

Goedert, Michel, Klug, Aaron, and Crowther, R.A. Tau protein, the paired helical filament and Alzheimer's disease. *Journal of Alzheimer's Disease*, 9:195–208, 2006.

Goodfellow, Ian, Warde-Farley, David, Lamblin, Pascal, Dumoulin, Vincent, Mirza, Mehdi, Pascanu, Razvan, Bergstra, James, Bastien, Frederic, and Bengio, Yoshua. Pylearn2: a machine learning research library. *arXiv*, 2013.

Gupta, A., Ayhan, M. S., and Maida, A. S. Natural image bases to represent neuroimaging data. *JMLR*, 28, 2013.

Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, Karayev, Sergey, Long, Jonathan, Girshick, Ros, Guadarrama, Sergio, and Darrell, Trevor. Caffe: Convolutional Architexture for Fast Feature Embedding. In *ACM MM 2014*, 2014.

Karpathy, Andrej. CS231n Convolutional Neural Networks for Visual Recognition. Online, 2015.

Kloppel, Stefan, Stonnington, Cynthia, Barnes, Josehpine, Chen, Frederick, Chu, Carlton, Good, Catriona, Mader, Irina, Mitchel, L. Anne, Patel, Ameet, Roberts, Catherine, Fox, Nick, Jack, Clifford, Ashburner, John, and Frackowiak, Richard. Accuracy of dementia diagnosisa direct comparison between radiologists and a computerized method. *Brain*, 131:2969–2974, 2008.

Krivhesky, Alex, Sutskever, Ilya, and Hinton, Geoffrey. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, 2012.

Langbaum, Jessica, Hendrix, Suzanne, Ayutyanont, Napatkamon, Chen, Kewei, Fleisher, Adam, Shah, Raj, Barnes, Lisa, Bennett, David, Tariot, Pierre, and Reiman, Eric. An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimer's & Dementia*, 10:666–674, 2014.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE*, 81(11), 1998.

LeCun, Yann, Kavukcuoglu, Koray, and Farabet, Clement. Convolutional Networks and Applications in Vision. In *Prov International Symposium on Circuits and Systems*, 2010.

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the Twenty-sixth International Conference on Machine Learning (ICML)*, 2009.

Ortiz, Andres, Gorriz, Juan, Ramirez, Javier, and Martinez-Murica, Francisco. Automatic ROI Selection in Structural Brain MRI Using SOM 3D Projection. *Plos One*, 2014.

Payan, A. and Montana, G. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *Computer Vision and Pattern Recognition*, 2: 355–362, 2015.

Penny, William, Friston, Karl, Ashburner, Juhn, Kiebel, Stefan, and Nichols, Thomas. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2006.

Plis, Sergey M., Hjelm, Devon R., Salakhutdinov, Ruslan, Allen, Elena A., Bockholt, Henry J., Long, Jeffrey D., Johnson, Hans J., Paulsen, Jane S., Turner, Jessica A., and Calhoun, Vince D. Deep learning for neuroimaging: A validation study. *Frontiers in Neuroscience*, 8:1–11, 2014.

Ramirez, Belen, Blazquez, Cristina, del Pulgar, Terea Gomez, Guzman, Manuel, and de Ceballos, Maria. Prevention of Alzheimer's Disease Pathology by Cannabinoids: Neuroprotection Mediated by Blockade of Microglial Activation. *Journal of Neuroscience*, 25 (8):1904–1913, 2005.

Reiman, Eric, Jangbaum, Jessica, Fleisher, Adam, Caselli, Richard, Chen, Kewei, Ayutyanont, Napatkamon, Quiroz, Yakeel, Kosik, Kenneth, Lopera, Francisco, and Tariot, Pierre. Alzheimers Prevention Initiative: A Plan to Accelerate the Evaluation of Presymptomatic Treatments. *J Alzheimers Dis*, 26, 2011.

Rolland, Yves, van Kan, Gabor Abellan, and Vellas, Bruno. Physical Activity and Alzheimer's Disease: From Prevention to Therapeutic Perspectives. *Journal of the American Medical Directors Association*, 9:390–405, 2008.

Shaffer, Jennifer L., Petrella, Jeffrey R., Sheldon, Forrest C., Choudhury, Kingshuk R., Calhoun, Vince D., Coleman, R.E, and Doraiswamy, P.M. Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology*, 266:583–591, 2013.

Suk, H. and Shen, D. Deep learning-based feature representation for ad/mci classification. *Med. Image. Comput. Comput. Assist. Interv.*, 16(2):583–590, 2013.

Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, and Rabinovich, Andrew. Going deeper with convolutions. *arXiv*, 2014.

Vincent, Pascal, Larochelle, Hugo, Bengio, Yoshua, and Manzagol, Pierre-Antoine. Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, 2008.

Ypsilantis, P., Siddique, M., Sohn, H., Davies, A., Cook, G., Goh, V., and Montana, G. Predicting response to neoadjuvant chemotherapy with pet imaging using convolutional neural networks. *Plos One*, 2015.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage.*, 55(3):856–867, 2011.

Zheng, Shuai, Jayasumana, Sadeep, Romera-Paredes, Bernardino, Vineet, Vibhaz, Su, Zhizhong, Du, Dalong, Huang, Chang, and Torr, Philip. Conditional Random Fields as Recurrent Neural Networks. In *International Conference on Computer Vision 2015*, 2015.