Classification of Alzheimer's Disease Based on White Matter Architecture

Tanya Glozman[1] and Rosemary Kim Le[2]

Abstract— **Alzheimers disease is the most common form of dementia in adults aged 65 or older. Although many studies have measured the effect of tissue degeneration for subcortical structures such as the hippocampi, amygdala, and the ventricles, little is known about the changes that occur in the architecture of the white matter during the course of this disease. The shape and properties of white-matter structures can be measured using Diffusion Tensor Imaging (DTI), a relatively new neuroimaging technique. Differences in these structures have been shown to be related to behavior, cognition, and neurological diseases [3]. In this project, we use machine learning tools to classify Alzheimer's patients and normal controls based on the architectural attributes of white matter tracts. The results are promising and demonstrate that the shape of white matter structures is useful in the quest to better understand the biomarkers and neural correlates of this disease.**

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive, degenerative brain disorder that leads to nerve cell death and tissue loss in the brain. It is the most common form of dementia is adults aged 65 and older. The worldwide prevalence of AD was reported 26.6 million in 2006 and is expected to rise to over 100 million by 2050 [1]. Many studies provide scientific evidence suggesting that AD progresses by selective degeneration of neuronal populations in several regions in the brain. Pathological structural changes in the brain occur many years before the symptoms of the disease are observed, at which stage significant neurodegeneration has already taken place. On the molecular scale, neuropathological hallmarks of AD are intraneuronal accumulation of neurofibrillary tangles, extracellular deposition of amyloid-beta protein as senile plaques, and massive neuronal death [7]. Definitive diagnosis of AD requires histopathologic confirmation of the presence of amyloid plaques and neurofibrillary tangles. In practice, clinical history and neuropsychological tests are used to confirm the diagnosis. *Identifying meaningful imaging-based biomarkers for early detection and diagnose of the disease are limited by lack of quantitative approaches to characterize how the disease affects the human brain at different stages during its degenerative advancement.*

In the past two decades, diffusion tensor imaging (DTI) has become one of the most important tools for neuroscience research and clinical practice [10] for its ability to model the integrity and structure of the brains white matter architecture in the living human brain. A white matter *fascicle*, also called *tract*, is a bundle of axon fibers that forms connections between different (often remote) areas of the brain. In general, DTI measures the motion of water molecules in the brain, which is free along the fibers and not across the fibers; this difference between parallel and perpendicular motion is called anisotropy. Fractional anisotropy is among the most widely used metrics in brain research [10] and is also commonly used in the study of Alzheimers patients [9]. Previous work done with Alzheimers disease and DTI primarily utilize statistical tests to demonstrate that there are differences in tissue measurements, such as fractional anisotropy, between these two populations. In this project, we explore whether the architectural features of the white matter may be informative in the classification of Alzheimer patients and normal controls. This paper will progress as follows: we will describe the data and our pre-processing pipeline. We will describe the features extracted from the data and the machine learning techniques that were applied for classification. We will then report the results and evaluate the performance of our method. We conclude with discusssion of the main insights and implications of this work and outline the most interesting future directions of this project.

II. DATASET AND PRE-PROCESSING PIPELINE

The data used in the preparation of this paper were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI database contains both T1-weighted MR imaging as well as DTI imaging data. Details about the acquisition protocol and the initial processing steps can be found in [6]. Many software tools for processing of DTI data are available through the Vista Lab at Stanford led by Professor Brian Wandell. Figure 1 outlines both the pre-processing steps and the classification framework. Below is a detailed explanation of the preprocessing steps performed for each subject:

- *AC-PC alignment of T1-weighted images.*
  High quality anatomical images (T1-weighted) serve as a spatial reference for the DTI data. This step is manual and involves defining three landmarks in the brain: the mid-sagittal plane, the anterior commissure, and the posterior commissure. This step resamples the data to align all images to the first one.
- *Conversion of the Axial DTI data from DICOM format to NIFTI format.*
  This step creates helper files that contain diffusion tensor directions and magnitudes.
- *mrTrix.*
  mrTrix [5] is a set of tools that performs diffusion-weighted MR white matter tractography in a manner

[1]T. Glozman is with Faculty of Electrical Engineering, Stanford University, Stanford, CA, 94305, USA. tanyagl@stanford.edu

[2]R. Le is with the Department of Psychology, Stanford University, Stanford, CA, 94305, USA. rosemary.le@stanford.edu

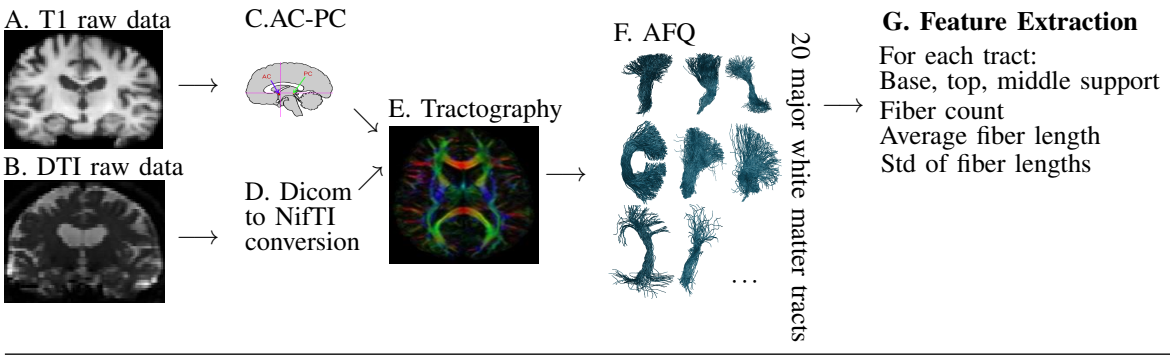

Fig. 1: **Overview of the classification framework.** The top panel outlines the pre-processing steps performed for every subject as well as the features extracted for the classification: (A) T1 raw data provides the anatomical reference. (B) DTI raw data. (C and D) canonical alighment and format conversion were performed on the T1-weighted data and the DTI data respectively. (E) Tractography algorithm estimates the direction and magnitude of water diffision at every voxel. (F) AFQ is run on the tractography output to find the 20 major white-matter tracts in the human brain. (G) Feature extraction: for each of the 20 tracts we extract the features listed above.
The lower panel outlines the classification procedure: first, we perform Feature Ranking on the training set based on the differences in the average population values for each feature. Next, classification is performed with approximately 40 testing examples and 140 training examples. Several machine learning methods were used. Following the classification, we evaluate our results as outlined above.

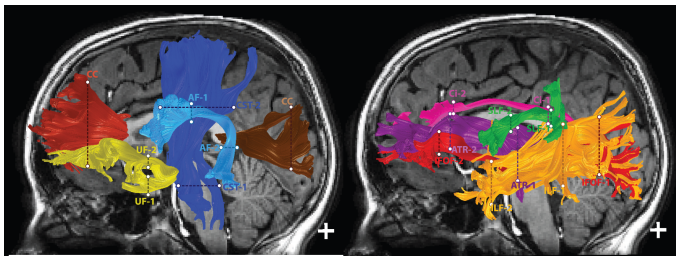

Fig. 2: The 20 tracts extracted from AFQ. The 120-dimensional feature vector was constructed from properties of these tracts. Picture courtesy of Yeatman, 2012 [11].

robust to crossing fibers using constrained spherical deconvolution and probabilistic streamlines.

- *Automatic Fiber Quantification (AFQ).* AFQ [11] is designed to automatically identify 20 major fiber tracts and quantify tissue properties along their trajectories. Figure 2 shows an example of 20 major white matter tracts extracted from AFQ with the T1-weighted MR image in the background for reference.
- *Fiber filtering and cleanup*. This step involves automatic cleaning and outlier removal of the generated fibers, using probabilistic and heuristic approaches.

The pre-processing time per scan is of the order of 2 hours. We pre-processed a total of 49 Alzheimer's Disease patients (AD) and 57 Normal Controls (NC) - a total of 239 scans - 100 AD scans and 139 NC scans. Many of these scans were longitudinal scans collected over a period of up to 3 years on the same subject (which is why we have more scans than subjects). In order to keep the training and testing set independent and uncorrelated, we randomly chose the subset of subjects to use for the testing set. Only one scan for these subjects was used for the testing set, the other longitudinal scans of the subject were not included in the train or test set. For the subjects chosen for the training set however, all longitudinal scans were used for training. As we elaborated in the milestone [4], the probabilistic nature of the tractography algorithms ensures that there will be a variation even between the longitudinal scans of the same subject. Using the longitudinal data in the training set allowed us to gain more training examples without jeopardizing the independence of the testing set. As a result of this data partitioning, for each experiment we have about 150 training examples and 40 testing examples, balanced across classes.

III. Features and Feature Ranking

For each scan, we constructed a 120-dimensional feature vector from the 20 extracted fibers tracts - 6 measures for each tract: number of fibers, the mean and standard deviation of all fiber lengths, and an estimate of "support" at the endpoints of each tract, as well as the "support" of the middle point of all fibers in a tract. "Support" was calculated as the volume of the cube that encapsulates the endpoints on each end and the middle. Intuitively, the endpoints "support"

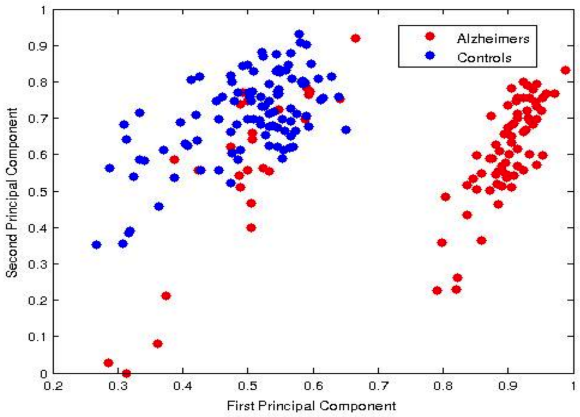

Fig. 3: Subjects' data projected onto the first two principal components. Separability can be observed between Alzheimer's patients (in red) and normal controls (in blue).

values provide us with the measure of the cortical projection of the fascicle, and these features turned out to be extremely important for the classification. All features were normalized to lie in the range $\{0,1\}$ according to the following formula:

$$\tilde{f} = (f - \min_i f_i)/(\max_j f_j - \min_i f_i)$$

where $i, j = 1, ..., n$, where n is the total number of examples.

We then ranked each feature according to how much its values differed between the populations, according to the following equation.

$$rank(feature) = \frac{|mean(AD) - mean(NC)|}{\sqrt{|std(AD) - std(NC)|}}$$

where AD stands for Alzheimer's Disease patients and NC stands for Normal Control subjects. This feature ranking was performed on the training set. Features were then sorted according to their ranking.

Figure 4 clearly shows that in the highest-ranked features (corresponding to features numbered 1-40), a separation is apparent between Alzheimer patients (in red) and normal controls (in blue). This was encouraging and suggested that a learning algorithm would be able to dinstinguish these classes with reasonable performance.

To further assess the separability of the data, we performed PCA (Principal Components Analysis) on the data. Figure 3 shows the data projected on the first and 2nd principal components. It is clear that the data is separable.

To find the optimal dimension $\tilde{n}$ of the feature vector for classification, we repeated the classification experiments using a growing set of features. $\tilde{n}$ was taken to be the number of features that resulted in the lowest testing error for the learned hypothesis. In our experiments, $\tilde{n}$ was 53 for both the RBF-kernel and linear-kernel SVMs and 36 for the logistic regression.

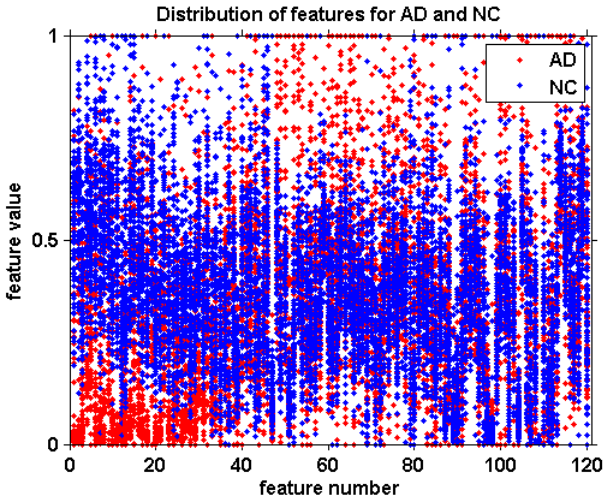

Fig. 4: The population distribution of features sorted from highest (at x=1) to lowest-ranked (at x=120) feature. Values for Alzheimer's patients are colored red and values for Normal controls are colored blue.

IV. Models

A. Logistic Regression

In logistic regression, our hypothesis is of the form

$$h_\theta(x) = \frac{1}{1 + \exp(-\theta^T x)}$$

where

$$\begin{cases} h_\theta(x) < 0 & \text{Alzheimer patient} \\ h_\theta(x) \geq 0 & \text{Normal Control} \end{cases}$$

The learned n-dimensional variable θ corresponds to the feature vector $x \in \mathbb{R}^n$ that corresponds to the n strongest features. The parameter θ is found by gradient ascent, where the update rule is as follows:

$$\theta_j := \theta j + \alpha(y^{(i)} - h_\theta(x^{(i)})x_j^{(i)}$$

To find the hypothesis that would result in minimal testing error, we iteratively calculate the hypothesis and the corresponding testing set error for feature dimensions $n = 1, 2, ...120$.

B. SVM

The support vector machine algorithm is calculated by minimizing

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} i^2 \xi_i$$

subject to

$$y^{(i)}(\omega^T x^{(i)} + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

for $i = 1, 2, ..., m$.

We used SVM with a linear and an RBF kernel. In the linear kernel,

$$K(\vec{x_i}, \vec{x_j}) = \vec{x_i}\vec{x_j}$$

In the RBF kernel,

$$K(\vec{x}_i,\vec{x}_j) = exp(\frac{-\|\vec{x}_i - \vec{x}_j\|^2)}{2\sigma^2}$$

As with logistic regression, we found the testing error for the classifier learned when using the n strongest features. We performed 5-fold cross validation to choose the optimal parameters for the classifier training. Each SVM experiment was repeated 20 times - the training and testing set were randomly chosen in each repeat, as explained in section II. The reported Training and Testing errors, Precision and Recall values are the average of these experiments. We also calculated the standard deviation of these experiment to make sure the results are consistent. Standard deviation was on the order of up to about 1%.

V. Results and Analysis

In Table I, we display training error, testing error, precision, and recall for the three classifiers. Precision p and recall r are defined as follows:

$$p = \frac{tp}{tp+fp}; r = \frac{tn}{tn+fp}$$

where tp = true positive; tn = true negative; fp = false positive; fn = false negative.

Figure 5 depicts the testing error achieved with an RBF kernel classifier as a function of number of features used. We see that the error achieved using just one feature is about 10%, and the error increases with adding the first few features. However, following this initial increase, the error drops to 7.93% at 53 features. It's interesting to note that once the testing error reaches it's minimal value, each additional feature just serves as noise in the data and "confuses" the classifier. This is consistent with figure 4 , where we clearly see that there is no difference between the two populations for features ranked 50 and up.

In order to evaluate our results, we show a comparison of the precision vs recall values of our best classifier (RBF-kernel SVM) to other works. A comparative study by Cuingnet et al. [2], evaluated the performance of several approaches using 509 subjects from the ADNI database. Note that none of these approaches used DTI data, as that data was not available at the time of the study. The classifiers used in these works were based on Gray-Matter data features such as volume of the hippocampi, cortex shrinking measures and voxel-based measurements. Additionally, as we are using a different dataset for our work, this comparison merely serves as a rough evaluation of the merit of our method.

VI. Discussion

Tractography algorithms are inherently uncertain and sensitive to choice of parameters [8]. Furthermore, data acquired from the ADNI database were collected from over 50 institutions with differing scan parameters. It is likely the results would be improved with standardized data collection, as well as with additional training data.

In this work we were interested in understanding the changes in the architecture of the white matter as a result

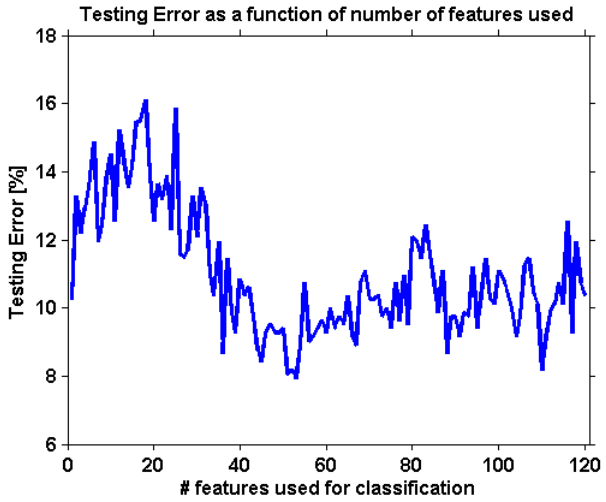

Fig. 5: Testing error achieved with an RBF kernel classifier as a function of number of features used. Note the initial increase in the error. Following this initial increase, the error drops to 7.93% at 53 features. It's interesting to note that once the testing error reaches it's minimal value, each additional feature just serves as noise in the data and "confuses" the classifier.

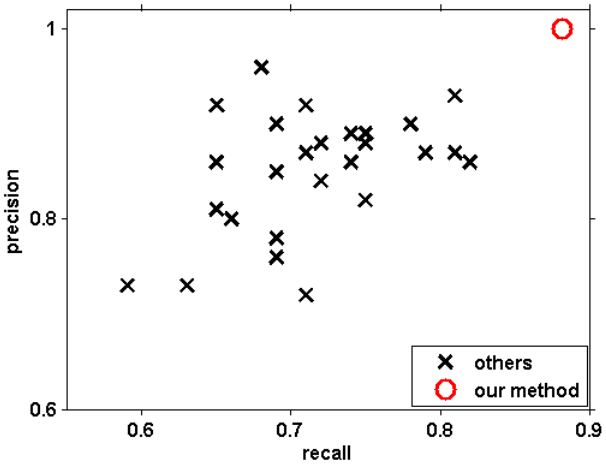

Fig. 6: Comparison to other works, based on [2]. Previous studies (as denoted by the x's) perform classification based on features calculated on the gray matter and other subcortical structures. Note, however that since we used DTI imaging data in this study, this comparison is not performed on the same dataset. At the date of the study of [2] DTI data was not yet available. Nevertheless, we believe that this comparison provides an estimate of the merit of our work and we find that using white matter properties as features results in a substantial improvement.

	Training Error (*)	Testing Error	Precision	Recall
SVM Linear	2.21	8.01	1	0.8705
SVM RBF Kernel	0.75	7.93	1	0.8816
Logistic Regression	0	7.32	0.9545	0.8947

TABLE I: Classification results for three models: SVM with Linear kernel, SVM with RBF kernel and Logistic Regression. About 150 training examples and 40 testing examples were used. (*) please note that the reported training error is very low. We believe the reason for this is that the training set contains correlated examples, since we used longitudinal scans data in this set to gain more examples. $\tilde{n} = 53$ for SVM and $\tilde{n} = 36$ for logistic regression.

of Alzheimer's disease, in addition to achieving good classification performance. We found it interesting that there is no statistically significant change in the number of fibers, and their average length following Alzheimer's disease. Even more so, we found that our top-ranked features were the "support" values for the major fascicles. As shown in figure 4, this feature showed separability between NCs and ADs, with ADs having a smaller value of "support". As the "support" can intuitively be considered to estimate the "volume" encapsulating the area of the cortical projection of a fascicle, we expect that as the cortex shrinks during the course of AD, this volume will decrease. It has been long known that tractography algorithms are especially unreliable near the cortex, thus we believe that our "support" feature is coarse enough to overcome this uncertainty, while still being accurate enough to be able to discriminate between AD and NC subjects.

In addition to constructing good features, feature normalization and ranking also proved to be extremely benefitial to the performance of the classifiers.

In the study of Alzheimer's disease, most machine learning techniques have been applied to measurements of gray matter and other subcortical structures. Since diffusion tensor imaging is a relatively recent neuroimaging technique, it is not yet widely used with regards to classification of Alzheimer's disease. The few works that do consider the white matter, focus on tissue properties, such as Fractional Anisotropy. To the best of our knowledge, this is the first work on classification of Alzheimers disease using architectural attributes of the white matter. We believe that shape and function in the human brain are strongly related, and are excited to continue exploring this direction.

VII. Conclusions

We show that the classification of Alzheimer's patients using white properties achieves state-of-the-art results. Diffusion tensor imaging, in conjunction with machine learning algorithms, has the ability to advance our understanding of Alzheimer's disease.

VIII. Future Directions

Including more complex geometric features that describe the architecture of the white matter may improve the classification results even further. We are also interested in understanding whether the features we constructed are useful as a screening tool for the prodromal stage of AD - the Mild Cognitive Impairment (MCI) disease. The ADNI dataset includes DTI imaging data for MCI patients along with the information whether or not these patients converted to Alzheimer's disease within a course of 18 months. Being able to identify the MCI-converters is of utter importance as this has the potential to aid clinicians in determining the course of treatment for these patients. Our next step will involve applying the method developed here on this additional data.

ACKNOWLEDGMENTS

The authors would like to thank the course staff for providing valuable feedback, Prof. Franco Pestilli and Dr. Hiromasa Takemura from the Vista Lab led by Prof. Brian Wandell for their advice and help with implementation of the image pre-processing pipeline.

References

[1] Ron Brookmeyer, Elizabeth Johnson, Kathryn Ziegler-Graham, and H. Michael Arrighi. Forecasting the global burden of alzheimers disease. *Alzheimer's amd Dementia*, 3(3):186 – 191, 2007.

[2] Remi Cuingnet, Emilie Gerardin, Jrme Tessieras, Guillaume Auzias, Stphane Lehricy, Marie-Odile Habert, Marie Chupin, Habib Benali, and Olivier Colliot. Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the {ADNI} database. *NeuroImage*, 56(2):766 – 781, 2011. Multivariate Decoding and Brain Reading.

[3] R. Douglas Fields. White matter in learning, cognition and psychiatric disorders. *Trends in Neuroscience*, 37(7):361–370, 2008.

[4] Tanya Glozman and Rosemary Kim Le. Classification of alzheimer's disease based on white matter attributes - cs229 milestone.

[5] Melbourne Australia J-D Tournier, Brain Research Institute. http://www.brain.org.au/software/.

[6] Clifford R. Jack, Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J. Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M. Dale, Joel P. Felmlee, Jeffrey L. Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A. Ward, Gregory J. Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[7] Yangling Mu and Fred H Gage. Adult hippocampal neurogenesis and its role in alzheimer's disease. *Molecular Neurodegeneration*, 6(85), 2011.

[8] Fillard P, Descoteaux M, Goh A, Gouttard S, Jeurissen B, Malcolm J, Ramirez-Manzanares A, Reisert M, Sakaie K, Tensaouti F, Yo T, Mangin JF, and Poupon C. Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom. *Neuroimage*, 56(1):220–234, 2011.

[9] Parente, D. Gasparetto, E. L. Cruz Jr, L. C. H. D. Domingues, R. C. Baptista, A. C. Carvalho, A. C. P., and Domingues R. C. Potential role of diffusion tensor mri in the differential diagnosis of mild cognitive impairment and alzheimer's disease. *American Journal of Roentgenology*, 190(5):1369–1374, 2008.

[10] Assaf Y. and Pasternak O. Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. *Journal of Molecular Neuroscience*, 34(1):51–61, 2008.

[11] Jason Yeatman. https://github.com/jyeatman/afq.