# Investigating the Relationship between Structural MRI and Personality Type

**Tayler Sindhu** tms598@nyu.edu
**Chaojie Zhang** cz2064@nyu.edu
*New York University*
*New York, NY 10016, USA*

## Abstract

Uncovering the biological basis of personality could have implications for the study of personality disorders. Existing work shows inconsistent relationships between regional brain volumes and personality. Contributing a deep learning perspective to the existing literature, we employ convolutional neural networks in an attempt to predict personality types, as measured by the NEO-FFI-3, from 3D brain magnetic resonance images from 1300 young adults from the Human Connectome Project. We use a pre-trained model developed based on prior work in Alzheimer's prediction, as well as models with custom architecture. We additionally explore class activation mapping and model compression. Our models were unable to predict personality type from imaging data, which may be a result of our model design, but alternatively may be secondary to a lack of a biological relationship between personality and brain volumes.

**Keywords:** Convolutional Neural Networks, Personality, MRI brain

## 1. Introduction

Uncovering biological insights into human personality could have important implications for the study of personality disorders and mental health. Several studies have explored the relationship between brain structure and personality traits. The majority of these studies employ voxel-based morphometry (VBM) (Cremers et al., 2011; Forsman et al., 2012; Kaasinen et al., 2005; Lu et al., 2014; Nostro et al., 2017), while others use volumetric data obtained from manual (Wright et al., 2007) or automated segmentation (Schutter et al., 2012; Yamasue et al., 2008).

Correlations are largely inconsistent across studies. For example, a VBM-based study found that extraversion was positively correlated with anterior cingulate cortex volume in males, and negatively correlated in females (Cremers et al., 2011). In contrast, others found this trait to be correlated with orbitofrontal cortex volume in both genders (DeYoung et al., 2010; Omura et al., 2005). Additionally, Koelsch et al. (2013) found neuroticism to be associated with increased left amygdala volume (Koelsch et al., 2013), while Cremers et al. (2011) found no such associations. Some authors postulate that these inconsistencies may be in part attributed to methodological variation, such as how variables, including gender, age and total grey matter volume are addressed (Hu et al., 2011).

To our knowledge, no study to date has used neural networks to predict personality traits from brain MRI scans. Such predictions could add to the existing body of work attempting to uncover the biological underpinnings of personality. Liu et al. (2019) demonstrated that convolutional neural networks can be used to classify subjects according to Alzheimer's

disease status based on MRI brain scans (Liu et al., 2019). We aim to contribute a deep learning approach to the existing body of research investigating the relationships between personality and regional brain volumes, in the hopes of contributing to mental health research. We postulate that neural networks can similarly be used to predict personality traits based on MRI brain scans in healthy, young adults, without the use of hand-crafted features.

## 2. Data

We used the Washington University-University of Minnesota Human Connectome Project Consortium 1200 Subjects Release, which includes T1-weighted, 3T MRI brain scans and personality data from 1,113 healthy adults between 22 and 35-years-old (Van Essen et al., 2012). MRI scans, acquired using the 3D MPRAGE protocol (Mugler III and Brookeman, 1990), are pre-processed (Glasser et al., 2013) format using FreeSurfer (Fischl, 2012) and FSL (Woolrich et al., 2009), which includes gradient distortion correction, brain extraction, bias field correction, and nonlinear registration to Montreal Neurological Institute space (Evans et al., 1993). Each image was originally $260x311x260$ voxels, with each voxel $0.7mm^3$.

Personality data is assessed using the the NEO Five-Factor Inventory-3 (NEO FFI-3), a validated, self-reported, 60-item tool which assesses "neuroticism", "extraversion", "openness", "agreeableness" and "conscientiousness" using five-point Likert scales (McCrae and Costa, 2010). We normalized raw personality scores into psychometric T-scores using the following equations per the NEO FFI-3 professional manual (McCrae and Costa, 2010):

$$\text{Z-score} = \frac{\text{Raw Score - Mean}}{\text{Standard Deviation}}; \;\; \text{T-score} = (\text{Z-score} * 10) + 50,$$

where mean and standard deviation are provided for adults greater than 21-years-old, stratified by sex. Subsequently, we binned scores into categories of "low", "average", and "high" for each of the five personality types. Scores less than or equal to 44 were considered low, scores between 45 and 55 were considered average, and scores greater than or equal to 56 were considered high. The distributions, which varied across personality types, are shown in Figure 1. For example, Only 15.7% of subjects were characterized as "low" in terms of conscientiousness.

## 3. Methods

We developed four 3-dimensional convolutional neural networks for our multi-class classification task of predicting each of the five personality types. We also used class activation mapping (CAM) with global average pooling to perform object localization, in an attempt to elucidate the brain regions related to specific personality traits (Zhou et al., 2016). Finally, we explored model compression to reduce the computational cost of our models. In all models, we employed cross entropy loss and stochastic gradient descent. We randomly split data into training (60%), validation(20%), and test (20%) tests. For each personality type, we used a distinct model, and calculated the classification accuracy and area under the curve of the receiver operating characteristic (AUROC) for each class. We used Pytorch

Figure 1: Distribution of binned scores across personality type. (From left to right: N: Neuroticism, E: Extraversion, O: Openness, A: Agreeableness, C: Conscientiousness.)


(Paszke et al., 2019) to build and evaluate models. We ran models on a single GPU. We used batch sizes of one, primarily due to memory constraints.

**Model 1:** We modified the best-performing model for Alzheimer's disease classification from Liu et al.'s public repository to accomodate our image sizes (Figure 2). We explored transfer learning, initializing the first 12 out of 15 layers using pre-trained weights without freezing of layers. We used trilinear interpolation to reduce image size. We tuned interpolation scale factors (0.3, 0.4, 0.5, 0.6,) and learning rates (0.00001, 0.0001, 0.001, 0.01), training each of these 16 combinations for 50 epochs.

**Model 2:** We used a custom architecture adapted from Liu et al.'s design, primarily tailored to accommodate our input size, with an additional convolutional block and different feature map sizes. We also used batch normalization instead of instance normalization (Figure 2) After training the model, we performed a compression simulation with Pytorch (Paszke et al., 2019). We explored the parameter distribution in the model, and replaced parameters below a threshold of 0.0258 with zero.

**Model 3:** We decreased the number of parameters by removing one of the two final linear layers in favor of an adaptive max pooling layer, and used CAM to allow the classification model to localize class-specific regions using the final feature maps before the global average pooling layer.(Zhou et al., 2016) (Figure 2).

**Model 4:** We attempted an additional custom architecture with random cropping, different kernel sizes, and residual connections, based on the structure of ResNet (He et al., 2016) and Inception (Szegedy et al., 2016) (Figure 2).

## 4. Results

For model 1, the hyperparameter combination resulting in the highest accuracy was an interpolation scale of 0.4, and a learning rate of 0.01 (Figure 3). For model 2, accuracy and average AUROC is shown across all personality types in Figure 4.

Figure 2: Model architectures. For model 1, outputs for an interpolation scale of 0.4 are shown. For model 4, Each (7,7,7) kernel is separated into three (1,1,7), (1,7,1), (7,1,1) kernels.

**Model 1**

| Block | Layer | Output Size |
|---|---|---|
| | Input | (1, 260, 311, 260) |
| | Interpolation(scale=0.4) | (1, 104, 124, 104) |
| 1 | Conv3d(1, 32, 1) | (32, 51, 61, 51) |
| | InstanceNorm3d(32) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=2, dilation=1) | |
| 2 | Conv3d(32, 256, 3, dilation=2) | (256, 23, 28, 23) |
| | InstanceNorm3d(256) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=2, dilation=1) | |
| 3 | Conv3d(256, 512, 5, dilation=2) | (512, 9, 11, 9) |
| | InstanceNorm3d(512) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=2, dilation=1) | |
| 4 | Conv3d(512, 512, 3, dilation=2) | (512, 1, 1, 1) |
| | InstanceNorm3d(512) | |
| | ReLU() | |
| | AdaptiveMaxPool3d((1,1,1)) | |
| FC1 | Linear(512, 1024) | (1024) |
| FC2 | Linear(1024, 3) | (3) |

**Model 2**

| Block | Layer | Output Size |
|---|---|---|
| | Input | (1, 260, 311, 260) |
| 1 | Conv3d(1, 8, 3, stride=1, padding=1) | (8, 86, 103, 86) |
| | BatchNorm3d(8) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 2 | Conv3d(8, 32, 3, stride=1, padding=1) | (32, 28, 34, 28) |
| | BatchNorm3d(32) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 3 | Conv3d(32, 64, 3, stride=1, padding=1) | (64, 9, 11, 9) |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 4 | Conv3d(64, 64, 3, stride=1, padding=1) | (64, 3, 3, 3) |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 5 | Conv3d(64, 64, 3, stride=1, padding=1) | (64, 1, 1, 1) |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| FC1 | Linear(64, 1000) | (1000) |
| FC2 | Linear(1000, 3) | (3) |

**Model 3**

| Block | Layer | Output Size |
|---|---|---|
| | Input | (1, 260, 311, 260) |
| 1 | Conv3d(1, 8, 3, stride=1, padding=1) | (8, 86, 103, 86) |
| | BatchNorm3d(8) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 2 | Conv3d(8, 32, 3, stride=1, padding=1) | (32, 28, 34, 28) |
| | BatchNorm3d(32) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 3 | Conv3d(32, 64, 3, stride=1, padding=1) | (64, 9, 11, 9) |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 4 | Conv3d(64, 64, 3, stride=1, padding=1) | (64, 4, 5, 4) |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=2) | |
| 5 | Conv3d(64, 128, 3, stride=1, padding=1) | (128, 4, 5, 4) |
| | BatchNorm3d(128) | |
| | ReLU() | |
| GAP | AdaptiveMaxPool3d((1,1,1)) | (128, 1, 1, 1) |
| FC | Linear(128, 3) | (3) |

**Model 4**

| Block | Layer | Output Size |
|---|---|---|
| | Input | (1, 260, 311, 260) |
| | Ramdom Crop | (1, 250, 300, 250) |
| 1 | Conv3d(1, 16, 3) | (16, 86, 103, 86) |
| | BatchNorm3d(16) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 2 | Conv3d(16, 16, 3) \| Conv3d(16, 16, 7) | (32, 27, 33, 27) |
| | Concat and Residual Connection | |
| | BatchNorm3d(32) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 3 | Conv3d(32, 32, 3) \| Conv3d(32, 32, 7) | (64, 9, 11, 9) |
| | Concat and Residual Connection | |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | MaxPool3d(kernel_size=3, stride=3) | |
| 4 | Conv3d(32, 32, 3) \| Conv3d(32, 32, 7) | (128, 1, 1, 1) |
| | Concat and Residual Connection | |
| | BatchNorm3d(64) | |
| | ReLU() | |
| | AdaptiveMaxPool3d((1,1,1)) | |
| FC | Linear(128, 3) | (3) |

Figure 3: Model 1 Predicting Neuroticism: Accuracy versus interpolation scale, stratified by learning rate.

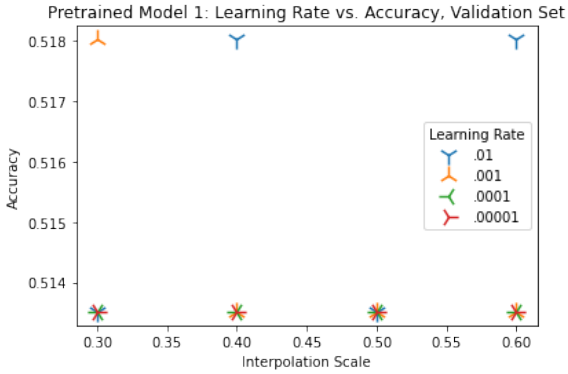



Figure 4: Model 2: Accuracy and area under receiver operating characteristic (AUROC) across predicted personality types.

| Model 2 | | | | | |
|---|---|---|---|---|---|
| **Predicted Personality** | **Neuroticism** | **Openness** | **Conscientiousness** | **Extraversion** | **Agreeableness** |
| **Average AUROC** | 0.50 | 0.54 | 0.48 | 0.50 | 0.51 |
| **Accuracy** | 41.00% | 38.74% | 38.74% | 40.09% | 38.29% |

Pruning of model 2 allowed for conversion to a sparse weight network. The original model contained 352,307 parameters, tightly clustered around zero (Supplemental Figure 1). After pruning, the number of parameters (excluding those in batch normalization layers) decreased by 79.41% to 72,737. We evaluated the compressed model predicting conscientiousness on the test set. Compared to the uncompressed model, the standard error of the class probabilities was 0.1011, and 4.05% of class predictions differed. Visualization of CAM (Model 3) is available in Supplemental Figure 2. For all models, accuracy and average AUROC is shown in Figure 5.

## 5. Discussion

Despite attempts to alter model architecture, tune hyperparameters, and employ transfer learning, our models were unable to predict personality based on brain region volumes. It is

Figure 5: Models 1-4: Accuracy and area under receiver operating characteristic (AUROC) across selected predicted personality types.

| **Model** | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
|---|---|---|---|---|
| **Predicted Personality** | Neuroticism | All | Agreeableness | Agreeableness |
| **Average AUROC** | 0.51 | 0.51 | 0.59 | 0.52 |
| **Accuracy** | 51.35% | 39.37% | 45.05% | 41.44% |

biologically plausible that structural imaging data alone does not contain the requisite data to predict personality type. This aligns with the inconsistent relationships found in existing work. It is also possible that our model design did not capture existing relationships.

Our first two models overfit the training data significantly. Our image pre-processing, which retained HCP's standard methods, varied significantly from Liu et al.'s approach. Although we used random cropping in model 4, combining Liu et al.'s image pre-processing methods with random cropping may have been beneficial. Additionally, Model 1 could benefit from exploring higher learning rates and increased batch sizes aided by model compression.

Model 2's custom architecture was able to meet memory requirements without downsampling. Model compression decreased computational requirements and resulted in only a minor change in output probabilities and in classification. CAM allowed for model visualization. Model 4, which was inspired by ResNet (He et al., 2016) and Inception (Szegedy et al., 2016), attempted to improve upon prior models by using random cropping, residual connections, and different kernel sizes, but did not result in significant improvement over other models.

## 6. Conclusion

Our models were unable to predict personality type from MRI brain images alone. This may reflect an underlying lack of biological relationship between the two entities, or that we did not employ the design choices necessary to build a predictive model. As some studies have suggested gender-specific differences between brain structure and personality (Nostro et al., 2017), we could explore the possibility of incorporating gender into our model. We could also obtain additional data, add variables such as age, gender, and total grey matter volume, or investigate alternate image preprocessing strategies, such as those which Liu et al. employed.

## 7. Contributions

Both teammates contributed equally, with Tayler focusing primarily on model 1, and Chaojie model 2, 3, and 4. The project was highly collaborative, with each partner discussing and helping each other with many aspects of the project.

## Acknowledgments

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

## 8. Appendix A

Figure Supplemental Figure 1: Model 2: Parameter distribution prior to compression

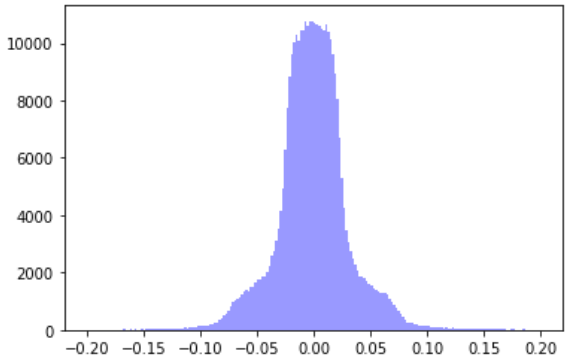
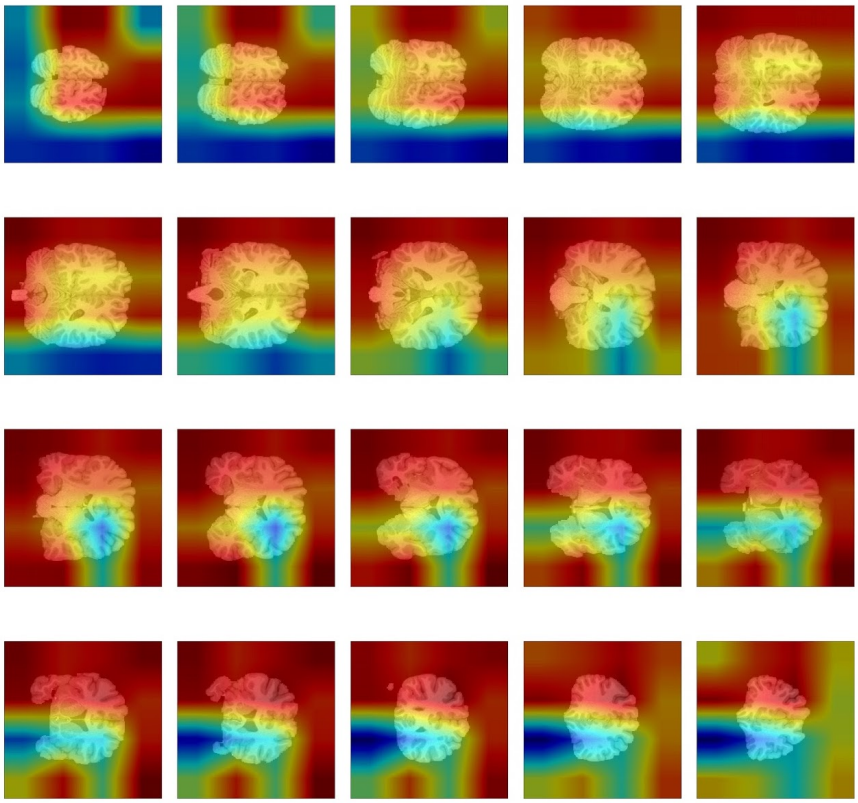



Figure Supplemental Figure 2: Model 3: Class activation mapping



## References

Henk Cremers, Marie-Jose van Tol, Karin Roelofs, Andre Aleman, Frans G Zitman, Mark A van Buchem, Dick J Veltman, and Nic JA van der Wee. Extraversion is linked to volume of the orbitofrontal cortex and amygdala. *PloS one*, 6(12), 2011.

Colin G DeYoung, Jacob B Hirsh, Matthew S Shane, Xenophon Papademetris, Nallakkandi Rajeevan, and Jeremy R Gray. Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological science*, 21(6):820–828, 2010.

Alan C Evans, D Louis Collins, SR Mills, ED Brown, RL Kelly, and Terry M Peters. 3d statistical neuroanatomical models from 305 mri volumes. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pages 1813–1817. IEEE, 1993.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Lea J Forsman, Örjan de Manzano, Anke Karabanov, Guy Madison, and Fredrik Ullén. Differences in regional brain volume related to the extraversion–introversion dimension—a voxel based morphometry study. *Neuroscience research*, 72(1):59–67, 2012.

Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Xiaochen Hu, Michael Erb, Hermann Ackermann, Jason A Martin, Wolfgang Grodd, and Susanne M Reiterer. Voxel-based morphometry studies of personality: Issue of statistical model specification—effect of nuisance covariates. *Neuroimage*, 54(3):1994–2005, 2011.

V Kaasinen, RP Maguire, T Kurki, A Brück, and JO Rinne. Mapping brain structure and personality in late adulthood. *Neuroimage*, 24(2):315–322, 2005.

S. Koelsch, S. Skouras, and S. Jentschke. Neural correlates of emotional personality: a structural and functional magnetic resonance imaging study. *PLoS One*, 8:e77196, 2013.

Sheng Liu, Chhavi Yadav, Carlos Fernandez-Granda, and Narges Razavian. On the design of convolutional neural networks for automatic detection of alzheimer's disease. *arXiv preprint arXiv:1911.03740*, 2019.

Fengmei Lu, Yajun Huo, Meiling Li, Heng Chen, Feng Liu, Yifeng Wang, Zhiliang Long, Xujun Duan, Jiang Zhang, Ling Zeng, et al. Relationship between personality and gray matter volume in healthy young adults: a voxel-based morphometric study. *PloS one*, 9 (2), 2014.

Robert R McCrae and Paul T Costa. *NEO Inventories For The NEO Personality Inventory-3 (NEO-PI-3), NEO Five-Factor Inventory-3 (NEO-FFI-3), NEO Personality Inventory-Revised (NEO PI-R): Professional Manual*. PAR, 2010.

John P Mugler III and James R Brookeman. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3d mp rage). *Magnetic resonance in medicine*, 15(1):152–157, 1990.

Alessandra D Nostro, Veronika I Müller, Andrew T Reid, and Simon B Eickhoff. Correlations between personality and brain structure: a crucial role of gender. *Cerebral cortex*, 27(7):3698–3712, 2017.

Kazufumi Omura, R Todd Constable, and Turhan Canli. Amygdala gray matter concentration is associated with extraversion and neuroticism. *Neuroreport*, 16(17):1905–1908, 2005.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Dennis JLG Schutter, P Cédric MP Koolschijn, Jiska S Peper, and Eveline A Crone. The cerebellum link to neuroticism: a volumetric mri association study in healthy volunteers. *PLoS One*, 7(5), 2012.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M Smith. Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45(1):S173–S186, 2009.

Christopher I Wright, Eric Feczko, Bradford Dickerson, and Danielle Williams. Neuroanatomical correlates of personality in the elderly. *Neuroimage*, 35(1):263–272, 2007.

Hidenori Yamasue, Osamu Abe, Motomu Suga, Haruyasu Yamada, Hideyuki Inoue, Mamoru Tochigi, Mark Rogers, Shigeki Aoki, Nobumasa Kato, and Kiyoto Kasai. Gender-common and-specific neuroanatomical basis of human anxiety-related personality traits. *Cerebral Cortex*, 18(1):46–52, 2008.

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.