

# ScholarWorks@GSU

## Brain ATLAS Driven Attention Models in ABCD Cognition Score Prediction

Item Type	Thesis
Authors	Deshpande, Sanket
Citation	Deshpande, Sanket (2025). "Brain ATLAS Driven Attention Models in ABCD Cognition Score Prediction." Thesis, Georgia State University. <a href="https://doi.org/10.57709/W3JX-4N16">https://doi.org/10.57709/W3JX-4N16</a>
DOI	<a href="https://doi.org/10.57709/W3JX-4N16">https://doi.org/10.57709/W3JX-4N16</a>
Download date	2026-04-13 10:32:07
Link to Item	<a href="https://hdl.handle.net/20.500.14694/15712">https://hdl.handle.net/20.500.14694/15712</a>

Brain ATLAS Driven Attention Models in ABCD Cognition Score Prediction

by

Sanket Deshpande

Under the Direction of Jingyu Liu, Ph.D.

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2025

## ABSTRACT

Emerging evidence points to the cerebellum as a key player in working memory and executive function, yet its structural role within the fronto-thalamo-cerebellar (FTC) circuitry remains underexamined. Leveraging the expansive ABCD dataset, this study evaluated the predictive capacity of FTC anatomy—specifically gray matter volume (GMV) and fractional anisotropy (FA)—in modeling working memory (2-back accuracy) and attention (0-back accuracy). Classical models such as Bayesian regression, SVR, and feedforward neural networks achieved a peak  $R^2$  of 0.062, with FTC-based predictions significantly outperforming fronto-parietal counterparts ( $p < 0.05$ ). Building on this, we developed a multi-stage deep learning framework incorporating 3D CNNs, self-attention blocks, and squeeze-and-excitation mechanisms. These models pushed predictive  $R^2$  to 0.079 while offering interpretable attention maps that consistently highlighted cerebellar subregions. Together, these results underscore the FTC circuit's robust involvement in working memory and demonstrate how deep learning can reveal spatially grounded, anatomically meaningful patterns from structural MRI.

**INDEX WORDS:** working memory, cerebellum, T1-weighted sMRI, gray matter volume, fractional anisotropy, ABCD Study, attention models, neural networks, machine learning, deep learning, CNN, attention maps, gradients, squeeze and excitation.

Copyright by  
Sanket Deshpande  
2025

Brain ATLAS Driven Attention Models in ABCD Cognition Score Prediction

by

Sanket Deshpande

Committee Chair: Jingyu Liu

Committee: Vince Calhoun,

Hemanth Venkateswara

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2025

## **DEDICATION**

I want to thank my family and friends for all their love and support. I appreciate the kindness and encouragement they gave me throughout my academic journey.

## ACKNOWLEDGEMENTS

I want to extend my gratitude to my advisor Dr. Jingyu Liu for all her support throughout my academic journey. I am grateful that I had the opportunity to work with her, as she has always encouraged me to strive for excellence. Her mentorship and guidance have been invaluable to my personal and professional growth. I would also like to thank my committee members Dr. Vince Calhoun and Dr. Hemanth Venkateswara for their expertise and contributions to my overall learning and growth at Georgia State University.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b>	<b>.....</b>	<b>V</b>
<b>LIST OF TABLES</b>	<b>.....</b>	<b>IX</b>
<b>LIST OF FIGURES</b>	<b>.....</b>	<b>X</b>
<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>1.1</b>	<b>ABCD Study .....</b>	<b>1</b>
<b>1.2</b>	<b>Thesis Outline.....</b>	<b>2</b>
<b>2</b>	<b>RELEVANT WORK .....</b>	<b>5</b>
<b>2.1</b>	<b>Imaging Modality Comparisons .....</b>	<b>5</b>
<b>2.2</b>	<b>Classical Regression and Machine Learning.....</b>	<b>5</b>
<b>2.3</b>	<b>Deep Learning Approaches for Cognition and ADHD Prediction.....</b>	<b>6</b>
<b>2.3.1</b>	<b><i>Functional MRI.....</i></b>	<b>6</b>
<b>2.3.2</b>	<b><i>T-1 Weighted Structural MRI.....</i></b>	<b>7</b>
<b>3</b>	<b>METHODOLOGY I.....</b>	<b>9</b>
<b>3.1</b>	<b>MRI Data Acquisition and Preprocessing .....</b>	<b>9</b>
<b>3.1.1</b>	<b><i>sMRI (Gray-Matter Volume) .....</i></b>	<b>9</b>
<b>3.1.2</b>	<b><i>dMRI (Fractional Anisotropy).....</i></b>	<b>9</b>
<b>3.2</b>	<b>ROI Feature Groups and Dataset Preparation .....</b>	<b>9</b>
<b>3.3</b>	<b>Technique Selection and Interpretation .....</b>	<b>11</b>
<b>3.3.1</b>	<b><i>Bayesian Ridge Regression .....</i></b>	<b>11</b>

3.3.2	<i>Support Vector Regression</i> .....	11
3.3.3	<i>Feedforward Neural Network</i> .....	11
4	<b>METHODOLOGY II</b> .....	13
4.1	<b>Pre-processing sMRI ROI Masks</b> .....	13
4.2	<b>K-Means Clustering of Masks</b> .....	14
4.3	<b>Atlas-Driven Patch Extraction</b> .....	15
4.4	<b>Baseline Concatenation Stage I Model</b> .....	17
4.5	<b>Single Head Self Attention Encoder</b> .....	19
4.6	<b>Squeeze-and-Excitation Path</b> .....	20
5	<b>RESULTS I</b> .....	23
5.1	<b>Performance of GMV vs FA features</b> .....	23
5.2	<b>Meta Analysis</b> .....	25
5.3	<b>ROI Interpretation and Discussion</b> .....	26
6	<b>RESULTS II</b> .....	29
6.1	<b>Training Results</b> .....	29
6.2	<b>Saliency Maps and Significant Volumes</b> .....	31
6.3	<b>Regions and Literature Discussion</b> .....	32
6.3.1	<i>Parietal and Thalamus</i> .....	32
6.3.2	<i>Cerebellar Lobules</i> .....	33
6.3.3	<i>Prefrontal Coretx</i> .....	34

<b>7</b>	<b>CONCLUSION .....</b>	<b>38</b>
	<b>APPENDICES .....</b>	<b>40</b>
	<b>Appendix A .....</b>	<b>40</b>

## LIST OF TABLES

Table 3-1 Distribution of ROIs Across Feature Groups and Imaging Modalities.....	10
Table 3-2 Participant Sample Allocation by Modality, Outcome Measure.....	10
Table 5-1 Performance of GMV Features .....	23
Table 5-2 Performance of FA features.....	24
Table 5-3 Meta-Analysis of feature group performances FTPC vs FTP vs FTC .....	25
Table 6-1 Baseline Concatentaion Training Performance .....	29
Table 6-2 Region Squeeze and Excitation Performacnce.....	30
Table 6-3Self Attention Encoder Performance.....	30
Table 6-4Anatomical regions and bilateral saliency-mask volumes).....	31
Table 7-1 Parameters of CNN1, CNN2, MLP Regressor .....	41

## LIST OF FIGURES

Figure 4-1 sMRI Masks of Different Regions .....	13
Figure 4-2 Clusters of the Individual Masks by Dimension x,y,z .....	14
Figure 4-3 Extraction of Tensors of two sizes .....	16
Figure 4-4 Architecture of Baseline Concatenation Stage I Training.....	18
Figure 4-5 Architecture of Self Attention Stage II Training.....	19
Figure 4-6 Original Squeeze and Excitation Block. ....	20
Figure 4-7 Architecture of Region Squeeze and Excitation Stage II Training .....	21
Figure 5-1 Top Regions from FTC and FTPC .....	27
Figure 6-1 Top Parietal Regions: Postcentral Gyrus and Thalamus.....	32
Figure 6-2 Top Cerebellar Lobules.....	34
Figure 6-3 Top Frontal Regions.....	36
Figure 7-1 Training Curves of Baseline, Self Attention, Squeeze and Excitation Modules .....	40

## 1 INTRODUCTION

Once thought to serve only motor coordination, the human cerebellum is now recognized as a key contributor to higher-order cognitive function [1]. A landmark observation by Schmahmann and colleagues in the late 1990s described a “cerebellar cognitive affective syndrome,” where cerebellar lesions led to impairments in executive function, attention, and emotional regulation [2]. Since then, accumulating evidence has shown that extensive cerebellar connections with cerebral cortex – particularly frontoparietal and limbic regions – enable the cerebellum to participate in cognition beyond motor control. In particular, attention and working memory, core components of executive function, have emerged as cognitive domains in which cerebellar involvement is consistently observed. There is a strong need for further evaluating and solidifying these claims with more empirical studies with different cohorts available at disposal. One such study that might elucidate this question is the ABCD Study, which not only is the most comprehensive study [3] of its kind, but a essential key used by research community for attention-deficit, ADHD and other neurodivergence clinical studies in a developing adolescent brain

### 1.1 ABCD Study

The Adolescent Brain Cognitive Development (ABCD) Study follows a cohort of nearly 12 000 children aged 9–10 years at enrollment, recruited from 21 sites across the United States to mirror national distributions of sex, race/ethnicity, and socioeconomic status [3]. Participants underwent a harmonized imaging protocol on 3 Tesla scanners, including high-resolution T1-weighted structural scans (1 mm isotropic), diffusion-weighted imaging (96 directions at 2 mm isotropic), and 10 minutes of resting-state gradient-echo EPI fMRI, alongside three task

paradigms: an emotional N-back working memory task (0-back vs. 2-back), a stop-signal inhibitory control task, and a monetary incentive delay reward task [4]. Cognitive performance was assessed using the NIH Toolbox Cognition Battery—which covers working memory, inhibitory control, processing speed, and language domains—to derive both domain-specific and composite scores [6]. Rigorous quality control procedures—including motion scrubbing with framewise displacement thresholds, automated artifact detection, and exclusion of behavioral outliers—yielded a final imaging sample of approximately 10 200 participants [4]. Key demographic covariates (sex: 51 % male, 49 % female; race/ethnicity: 21 % Black, 18 % Hispanic/Latino, 6 % Asian, 54 % White, 1 % Other; parental education; household income) were systematically recorded to control for potential confounds [5]. This rich phenotype, multimodal dataset provides an unparalleled resource for probing how structural and functional brain development underpins emerging cognitive abilities during early adolescence.

## **1.2 Thesis Outline**

There has been a lot of focus on using fMRI scans within the ABCD data to predict and understand cognition scores. This study takes a different modality into focus. Project 1 applies classical ROI-based analyses to T1-weighted structural scans—extracting mean gray-matter volumes from fronto-thalamo-parietal-cerebellar circuits—and predicts 4 metrics. 2-back working-memory, 0-back attention-vigilance, CBCL Assessment score, 1<sup>st</sup> PCA of CBCL and KSAD assessment scores using traditional machine learning and statistical methods. The study also does a deep dive into the different interpretability options within the methods to come up with important scores to rank and evaluate the ROIs. Project 2 then builds on those static summaries by hyper-focusing on t1-weighted sMRI scans and 2-back working memory deriving

data-driven feature embeddings from each region, integrating them into custom attention-based architectures that both boost predictive performance and yield interpretable maps of cerebellar involvement in working memory. This project explores state of the art deep learning methods available to test their efficacy and convergence to the findings of the first project , or diversions if any. Together, these studies meld traditional ROI approaches with modern embedding and attention mechanisms to deliver a multi-scale portrait of how cerebellar circuitry supports emerging executive functions in adolescence.

## REFERENCES

- [1] Brissenden JA, Somers DC. Cortico-cerebellar networks for visual attention and working memory. *Curr Opin Psychol*. 2019 Oct;29:239-247. doi: 10.1016/j.copsyc.2019.05.003. Epub 2019 May 21. PMID: 31202085; PMCID: PMC7256875.
- [2] Zhang P, Duan L, Ou Y, Ling Q, Cao L, Qian H, Zhang J, Wang J, Yuan X. The cerebellum and cognitive neural networks. *Front Hum Neurosci*. 2023 Jul 28;17:1197459. doi: 10.3389/fnhum.2023.1197459. PMID: 37576472; PMCID: PMC10416251.
- [3] Garavan H, Bartsch H, Conway K, DeCastri A, Goldstein RZ, Heeringa S, et al. Recruiting the ABCD sample: Design considerations and procedures. *Dev Cogn Neurosci*. 2018;32:16–22. doi:10.1016/j.dcn.2018.04.004.
- [4] Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al.; ABCD Imaging Acquisition Workgroup. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. 2018;32:43–54. doi:10.1016/j.dcn.2018.03.001.
- [5] Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci*. 2018;32:55–66. doi:10.1016/j.dcn.2017.10.010.
- [6] Luciana M, Bjork JM, Nagel BJ, Barch DM, Gonzalez R, Nixon SJ, Banich MT. Adolescent neurocognitive development and impacts of substance use: Overview of the ABCD baseline neurocognition battery. *Dev Cogn Neurosci*. 2018;32:67–79. doi:10.1016/j.dcn.2018.02.006.

## 2 RELEVANT WORK

### 2.1 Imaging Modality Comparisons

Research in large cohorts has increasingly contrasted structural versus functional MRI features for predicting cognitive abilities. For example, Dhamala et al. [1] found that in young adults, whole-brain functional connectomes were significantly more predictive of individual cognitive test scores than diffusion-derived structural connectomes—and that combining both modalities yielded no further gain. In developmental samples, Corriveau et al. [2] similarly reported superior performance from functional connectivity features in adolescents, with structural metrics contributing mainly to crystallized domains but underperforming for dynamic executive tasks.

Nevertheless, structural MRI (e.g., regional volumes, cortical thickness) retains predictive value for traits with clear anatomical substrates and offers complementary information about brain maturation. Indeed, Kardan et al. [3] demonstrated that functional connectivity patterns predictive of working memory reorganize from childhood to adulthood, as adult-derived models generalize poorly to youth—highlighting that modality importance shifts with development.

### 2.2 Classical Regression and Machine Learning

Early predictive studies in ABCD and similar cohorts applied traditional regression and shallow learners to extracted ROI features. Owens et al. [4] used elastic-net regression on task-evoked fMRI activations and structural morphology in ~11,000 children to predict ADHD symptoms, finding that distributed fronto-striatal-cerebellar patterns explained a small but significant portion of variance. Duffy et al. [5] trained a Poisson elastic-net on resting-state

connectivity in ~8,000 youth, identifying reduced default-mode–dorsal-attention anticorrelation as predictive of attention problems. These examples illustrate the feasibility—and limits—of classical models (linear regression, SVR, random forests) applied to summarized MRI metrics under rigorous cross-validation

## **2.3 Deep Learning Approaches for Cognition and ADHD Prediction**

Large-scale datasets have enabled deep learning methods—CNNs, graph nets, and Transformers—to automatically extract predictive features from high-dimensional MRI data, often surpassing traditional models and offering built-in interpretability via attention weights or saliency maps.

### ***2.3.1 Functional MRI***

Graph-based and attention-augmented deep networks have delivered substantial gains in modeling functional MRI. Zhu et al. [6] proposed “Multi Band Brain Net,” which decomposes resting-state and task BOLD signals into multiple frequency bands, constructs frequency-specific connectomes, and applies transformer-style self-attention to each band. In both ABCD and UK Biobank cohorts, this approach improved cognitive and psychiatric prediction accuracy by ~30 % over prior graph methods and consistently highlighted cerebellar-cortical circuits as key biomarkers. Lee et al. [7] demonstrated that representations learned on ABCD task fMRI for a general “p-factor” of psychopathology generalize across cultures: after fine-tuning on independent adolescent samples, the pretrained model achieved significantly higher correlations with working-memory and attention scores than classical kernel models, with attention maps revealing frontoparietal consistency and cohort-specific network contributions. More recently, Li et al. [8] introduced MHNet, which fuses graph neural networks for global connectivity with 3D

convolutions on local volumetric summaries. Applied to ADHD and ASD cohorts, MHNet outperformed standalone CNN and GNN baselines and identified cerebellar-cortical connectivity patterns that differentiated symptom severity, underscoring the value of hybrid attention models in functional connectomics.

### ***2.3.2 T-1 Weighted Structural MRI***

Deep learning on structural MRI has yielded incremental yet meaningful improvements over classical morphometric analyses. Mihalik et al. [9] combined voxel-wise 3D CNNs with kernel ridge regression on ABCD gray-matter volumes, winning the Neurocognitive Prediction Challenge by attaining the lowest mean squared error for fluid intelligence. Zhao et al. [10] further showed that end-to-end 3D CNNs on unsegmented T1 volumes outperform classical baselines, with learned feature maps correlating strongly with fluid intelligence and localizing to frontal and parietal cortices in line with the Parieto-Frontal Integration Theory. Chen et al. [11] adopted a two-stage pipeline—segmenting ROIs with a 3D U-Net then regressing on extracted sub volumes via CNN—which improved interpretability by isolating region-specific contributions. Finally, Valverde et al. [12] trained an autoencoder on 148 Free Surfer ROI features from over 11,000 ABCD participants; the resulting embeddings, when fed to a simple MLP, predicted fluid intelligence with  $r \approx 0.21$ , and SHAP analyses revealed sex-specific volumetric contributions in frontal and temporal regions. Despite these advances, predictive  $R^2$  typically remains below 0.10, suggesting structural anatomy alone may limit predictability without multimodal or finer-grained features

## REFERENCES

- [1] Dhamala, E., Zhang, P., Ryali, S., & Menon, V. Comparative Predictive Power of Structural, Diffusion, and Functional MRI for Cognition. *Cerebral Cortex*, 32(4), 1687–1703 (2022).
- [2] Corriveau, A., Li, X., Ho, T. C., et al. Connectivity Stability Correlates with Attention and Working Memory Across Samples. *Human Brain Mapping*, 44(2), 405–417 (2023).
- [3] Kardan, O., Henderson, J., Dalton, M., et al. Developmental Reorganization of Attention and Working Memory Networks in Children and Adults. *Nature Communications*, 13, 4129 (2022).
- [4] Owens, E., Dougherty, L. R., O'Brien, K. M., et al. Elastic Net Prediction of ADHD Symptoms from Task fMRI in the ABCD Study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(7), 634–643 (2021).
- [5] Duffy, B. A., Vanderwal, T., Castellanos, F. X., et al. Resting State Connectivity and CBCL Attention Problems in ABCD Children: A Poisson Elastic Net Approach. *Journal of Child Psychology and Psychiatry*, 64(6), 731–742 (2023).
- [6] Zhu, X., Liu, Q., Wang, Y., et al. Multi Band Brain Net: A Transformer Based Framework for Frequency Resolved Spatiotemporal Dynamics in fMRI Predicting Cognitive and Psychiatric Outcomes. *NeuroImage*, 310, 119772 (2025).
- [7] Lee, H., Kim, J., Park, S., et al. Transfer Learning of Deep Neural Networks Pretrained on ABCD Task Based fMRI for Cross Cultural Prediction of Working Memory and ADHD Symptoms. *Human Brain Mapping*, 46(8), 1754–1768 (2025).
- [8] Li, Y., Smith, R., Zhou, X., et al. MHNet: Multi View High Order Network Integrating CNN and GNN for Resting State fMRI Analysis in ADHD and ASD. *Nature Machine Intelligence*, 7(3), 287–298 (2025).
- [9] Mihalik, N., Ban, J., Sabuncu, M. R., et al. ABCD Neurocognitive Prediction Challenge: Voxel Wise CNN and Kernel Ridge Regression for Fluid Intelligence from T1 Weighted MRI. *Proceedings of the ABCD Challenge*, 1–9 (2019).
- [10] Zhao, W., Cai, X., Zhang, H., et al. 3D Convolutional Neural Networks for Prediction of Fluid Intelligence from Structural MRI. *Neuroinformatics*, 19(2), 211–222 (2021).
- [11] Chen, B., Wang, L., Li, C., et al. A Two Stage 3D U Net and CNN Regression Pipeline for Fluid Intelligence Prediction from T1 Weighted MRI. *Medical Image Analysis*, 67, 101821 (2021).
- [12] Valverde, G., Oliver, A., & Lladó, X. Predicting Fluid Intelligence in Adolescents from MRI Derived Morphometric Features Using Neural Networks. *Human Brain Mapping*, 40(2), 476–486 (2019).

## 3 METHODOLOGY I

### 3.1 MRI Data Acquisition and Preprocessing

#### 3.1.1 *sMRI (Gray-Matter Volume)*

T1-weighted scans were processed in MATLAB 2023b using SPM12's DARTEL toolbox. Steps included unified bias-field correction, tissue segmentation, and nonlinear registration to template space. Resulting gray-matter volume (GMV) maps were resliced to 1.5 mm isotropic voxels and smoothed with a 3 mm FWHM Gaussian kernel.

#### 3.1.2 *dMRI (Fractional Anisotropy)*

Diffusion images underwent eddy-current, motion, and EPI-distortion correction via FSL's eddy (with outlier replacement), followed by tensor fitting with dti-fit to compute fractional anisotropy (FA) maps. FA volumes were nonlinearly registered to MNI space using ANTs and resliced to 1 mm isotropic voxels.

### 3.2 ROI Feature Groups and Dataset Preparation

We defined three anatomically motivated circuits that capture distinct fronto-thalamo-parietal-cerebellar pathways (see Table 3.1). ROIs were drawn from the Schaefer cortical atlas (frontal and parietal cortices), the Seitzman subcortical atlas (thalamic nuclei), the SUIT hierarchical cerebellar atlas, and an AHEAD-derived red-nucleus partition. Automated QC pipelines excluded scans failing preprocessing. These cleaned feature matrices underlie all subsequent nested cross-validation and meta-analysis procedures

Table 3-1 Distribution of ROIs Across Feature Groups and Imaging Modalities. *This table summarizes the number of structural (GMV) and diffusion (FA) ROIs included in each anatomically defined circuit.*

<b>Circuit</b>	<b>sMRI (GMV) ROIs</b>	<b>dMRI (FA) ROIs</b>	<b>Description</b>
FTPC	335	341	Fronto-Thalamo-Parietal- Cerebellar
FTP	304	310	Fronto-Thalamo-Parietal
FTC	211	213	Fronto-Thalamo-Cerebellar

To predict individual 2-back (c2b) working-memory accuracy from these ROI features, data were partitioned according to a nested  $5 \times 5$  CV framework. Table 3.2 details the split of samples across targets and validation folds.

Table 3-2 Participant Sample Allocation by Modality, Outcome Measure, and Cross-Validation Fold *Outlines the total number of subjects and their division into outer test, outer dev(training), and inner validation sets for each modality and target.*

<b>Modality</b>	<b>Outcome</b>	<b>Total N</b>	<b>Outer Test N</b>	<b>Outer Dev N</b>	<b>Inner Val N</b>
sMRI	c0b	10 299	2 059	8 239	1 648
sMRI	c2b	10 299	2 059	8 239	1 648
sMRI	cbcl	10 537	2 107	8 429	1 686
sMRI	pc1	6 928	1 385	5 543	1 109
dMRI	c0b	8 299	1 660	6 639	1 328
dMRI	c2b	8 299	1 660	6 639	1 328
dMRI	cbcl	8 475	1 695	6 780	1 356
dMRI	pc1	5 869	1 108	4 761	952

### 3.3 Technique Selection and Interpretation

#### 3.3.1 Bayesian Ridge Regression

Ridge regression was cast in a Bayesian framework by imposing zero-mean Gaussian priors on both the regression weights and the noise precision. The prior precision ( $\alpha_1$ ) and noise precision ( $\alpha_2$ ) are estimated concurrently through evidence maximization, yielding principled uncertainty quantification for all parameters [1]. Hyperparameter tuning was performed using Bayesian Ridge, with  $\alpha_1$  and  $\alpha_2$  sampled on a log-uniform grid from  $10^{-6}$  to  $10^2$ . Feature relevance was assessed by the absolute magnitude of each ROI's posterior mean weight, with larger values indicating greater model influence under the learned Gaussian prior.

#### 3.3.2 Support Vector Regression

An  $\epsilon$ -insensitive support vector regression employing a radial-basis-function kernel accommodated potential nonlinearity in the ROI–behavior relationship [2]. The regularization parameter ( $C$ ) and kernel width ( $\gamma$ ) were optimized via nested cross-validation, exploring  $C \in [10^{-3}, 10^3]$  and  $\gamma \in [10^{-4}, 10^1]$  on a logarithmic scale. Unbiased estimates of each feature's contribution were obtained through permutation importance: shuffling the values of a single ROI in the held-out test set and measuring the resulting drop in  $R^2$ .

#### 3.3.3 Feedforward Neural Network

A multilayer perceptron comprising two hidden layers with Rectified Linear Unit activations and dropout regularization was trained using the Adam optimizer [3,4]. Hyperparameter selection involved grid searches over hidden-unit counts (Layer 1: {64, 128, 256}; Layer 2: {32, 64, 128}), dropout probabilities (0.1–0.5), and learning rates ( $10^{-4}$ – $10^{-2}$ ). Early stopping based on inner-fold validation loss prevented overfitting. Input-feature saliency

was quantified by backpropagating the network's output gradient to each ROI, thereby estimating the sensitivity of predictions to infinitesimal feature perturbations [5].

## REFERENCES

- 1) Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).
- 2) Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Statistical Computing* **14**, 199–222 (2004).
- 3) Srivastava, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**, 1929–1958 (2014).
- 4) Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. in *Proc. 3rd Int. Conf. Learn. Represent.* (2015).
- 5) Simonyan, K., Vedaldi, A. & Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv:1312.6034* (2014).

## 4 METHODOLOGY II

### 4.1 Pre-processing sMRI ROI Masks

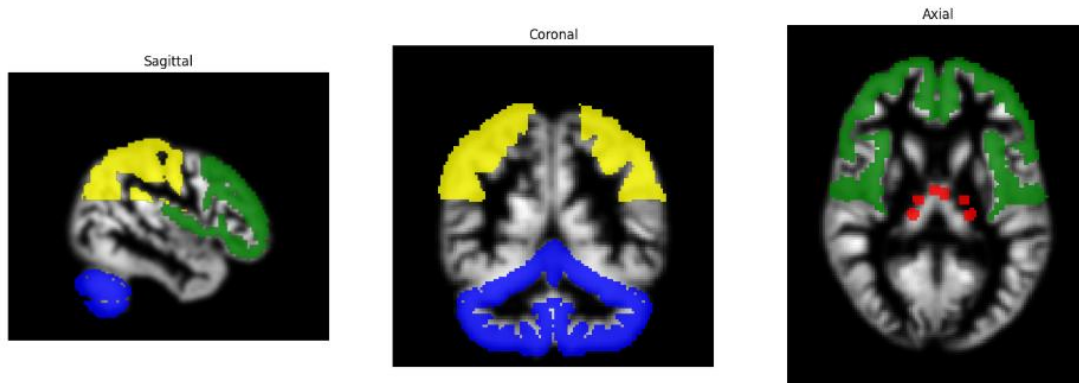


Figure 4-1 sMRI Masks of Different Regions

This figure visualizes the spatial distribution of anatomical regions of interest (ROIs) across sagittal, coronal, and axial planes using color-coded binary masks derived from standardized brain atlases. Green indicates frontal lobe regions, yellow marks parietal cortex, blue denotes cerebellar lobules, and red highlights thalamic nuclei, each internally composed of distinct subregions..

In preparation for region-aware analyses, atlas-based ROI definitions were converted into individual binary masks and accompanied by a comprehensive metadata table via a two-stage pipeline:

- 1. Label extraction and voxel enumeration.**

Each anatomical ROI volume (e.g., frontal, cerebellum, red nucleus, thalamus, parietal) was loaded from its NIfTI file.(See Figure 4-1) Unique integer labels were treated as subregion identifiers, and for each label all nonzero voxels were enumerated to record their 3D indices and total counts.

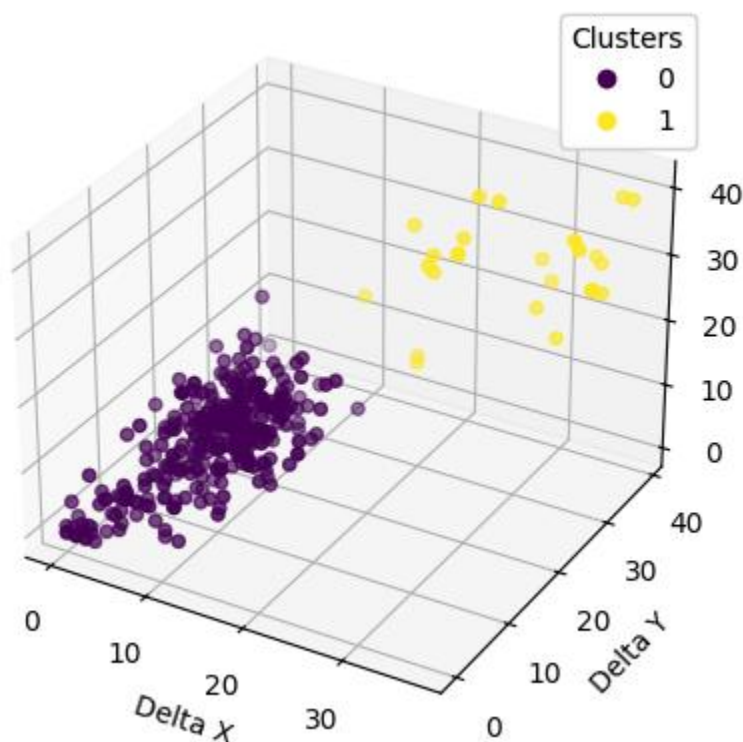
- 2. Mask generation and spatial metadata.**

For every identified label, a binary mask was constructed within the original  $121 \times 145 \times$

121 grid, setting voxels of that subregion to 1 and all others to 0. From each mask, the spatial extent— $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$ , defined by the minimum and maximum occupied coordinates—was derived, and the axis-aligned centroid of that bounding box was calculated.

The resulting per-region masks and associated metadata (voxel counts, extents, centroids) formed the basis for subsequent tensor extraction and attention-based modeling.

## 4.2 K-Means Clustering of Masks



*Figure 4-2 Clusters of the Individual Masks by Dimension  $x,y,z$*   
 This 3D scatter plot depicts the outcome of K-means clustering ( $K = 2$ ) applied to the spatial extents of individual sMRI-derived region masks, where each point represents a subregion characterized by its bounding-box dimensions along sagittal ( $\Delta X$ ), coronal ( $\Delta Y$ ), and axial ( $\Delta Z$ ) axes. Cluster 0 (dark purple) encompasses compact subcortical structures such as thalamic and cerebellar nuclei, while Cluster 1 (yellow) aggregates more spatially extensive cortical parcels.

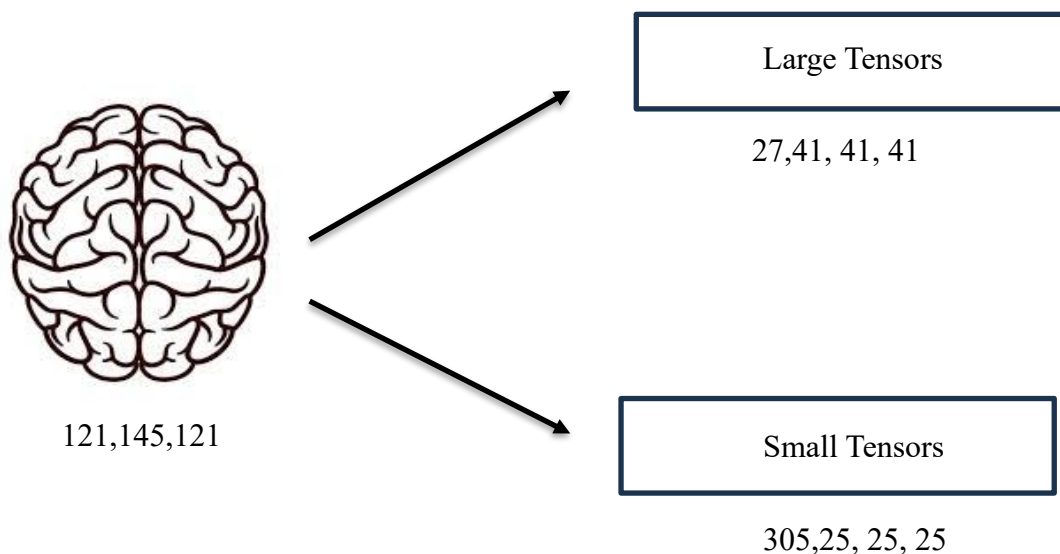
Each atlas-derived subregion was first summarized by its spatial extent— $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$ —defined as the range of nonzero voxels along the sagittal, coronal, and axial axes, respectively. To uncover latent morphological classes, K-means clustering ( $K = 2$ ) was applied to the resulting  $N \times 3$  feature matrix, where  $N$  is the number of subregions. (See Figure 4-2) Cluster assignments were visualized in a three-dimensional scatter plot of  $(\Delta X, \Delta Y, \Delta Z)$  to assess separation.

For each cluster, the following descriptive statistics were computed:

- Range (min–max) and mean of  $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$
- Number of subregions contained

This unsupervised partitioning consistently produced two groups: one characterized by larger extents—corresponding primarily to cortical parcels—and the other by smaller extents, representing subcortical nuclei. Incorporating these morphological classes into subsequent normalization and feature-selection steps ensures that downstream models appropriately account for region size and shape variability.

### 4.3 Atlas-Driven Patch Extraction



*Figure 4-3 Extraction of Tensors of two sizes*

*This schematic depicts the pipeline for generating standardized volumetric patches around atlas-defined brain regions from structural MRI scans. In Stage I, T1-weighted images (shape:  $1 \times 121 \times 145 \times 121$ ) are preprocessed via skull-stripping, bias correction, and intensity normalization. Precomputed ROI centroids—categorized into “small” and “large” masks based on their spatial extent—guide the patch extraction in Stage II:  $25^3$ -voxel patches for smaller subcortical regions and  $41^3$ -voxel patches for larger cortical parcels*

Figure 4-3 illustrates the two-stage, atlas-driven pipeline devised to separate volumetric patch extraction from downstream regression. In Stage I, each subject’s bias-corrected, skull-stripped T1 image ( $1 \times 121 \times 145 \times 121$  voxels) was loaded and intensity-normalized to zero mean and unit variance in PyTorch. Atlas-defined ROI centroids—pre-clustered into “small” and “large” groups based on their  $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$  extents—served as kernel centers for sampling.

In Stage II, cubic patches were extracted around each centroid using boundary-aware, zero-padded slicing:  $25^3$ -voxel kernels for smaller ROIs and  $41^3$ -voxel kernels for larger ROIs. As patches were generated, they were assembled into two tensors (cnn1\_data for small, cnn2\_data for large) alongside the normalized 2-back scores. These tensors were serialized to disk as cnn1.pt, cnn2.pt, and c2b.pt files within subject-specific directories.

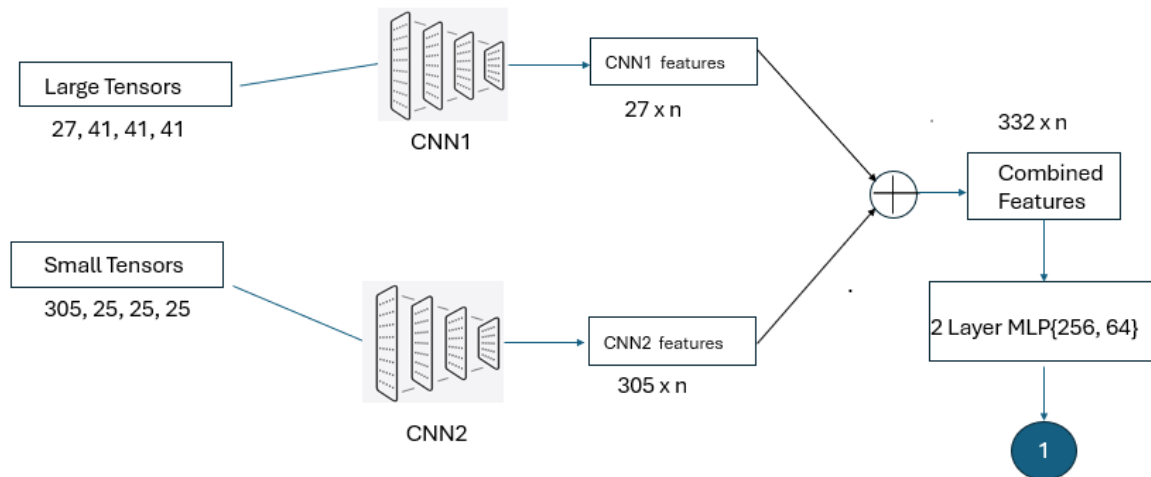
To maximize throughput on a multi-GPU cluster, the cohort was evenly split into N subsets, each processed in parallel via Python’s multiprocessing (spawn context). Within each GPU-bound worker, the MRI Dataset class performed on-the-fly patch extraction before writing the resulting files. This design overlaps I/O with computation and produces standardized, on-disk regional feature representations for reproducible training.

Two sampling schemas were evaluated. In sparse sampling, voxels outside the target ROI within each patch are zeroed out, reducing dimensionality and inter-parcel covariance but omitting boundary gradients. In dense sampling, all voxels within the cubic window are retained, preserving spatial continuity and texture at ROI interfaces at the cost of increased computational overhead

#### **4.4 Baseline Concatenation Stage I Model**

An end-to-end baseline regression architecture was constructed to simultaneously leverage information from both small and large ROI patches. (See Figure 4-4) In this dual-stream design, each subject's set of 27 "large" patches ( $41^3$  voxels each) is processed by a 3D-CNN encoder (CNN1), while the 305 "small" patches ( $25^3$  voxels each) are fed into a separate 3D-CNN encoder (CNN2). Each encoder applies a series of volumetric convolution and pooling layers to produce an n-dimensional feature vector for each region, yielding two latent embedding matrices of size  $27 \times n$  and  $305 \times n$ , respectively.

These per-region embeddings are concatenated along the region axis to form a combined feature map of dimension  $332 \times n$ , which is then flattened and input to a two-layer multilayer perceptron (MLP) with hidden layer sizes of 256 and 64. ReLU nonlinearities and dropout regularization are applied between layers, and a final linear neuron produces the working-memory score prediction. Training minimizes the mean squared error between predicted and actual 2-back accuracies, using the Adam optimizer with a carefully tuned learning rate and weight-decay schedule. Model selection relies on early stopping based on validation loss, and performance is reported in terms of  $R^2$  and Pearson correlation on an independent test set.



*Figure 4-4 Architecture of Baseline Concatenation Stage I Training*

*This figure illustrates an end-to-end regression pipeline that jointly integrates spatial embeddings from small and large brain ROIs. Large volumetric patches (27 regions,  $41 \times 41 \times 41$  voxels each) are processed by a dedicated 3D-CNN encoder (CNN1), while a separate 3D-CNN (CNN2) encodes the 305 smaller regions ( $25 \times 25 \times 25$  voxels). Each CNN outputs an  $n$ -dimensional feature vector per region, forming two latent matrices ( $27 \times n$  and  $305 \times n$ ), which are concatenated to produce a unified  $332 \times n$  representation.*

## 4.5 Single Head Self Attention Encoder

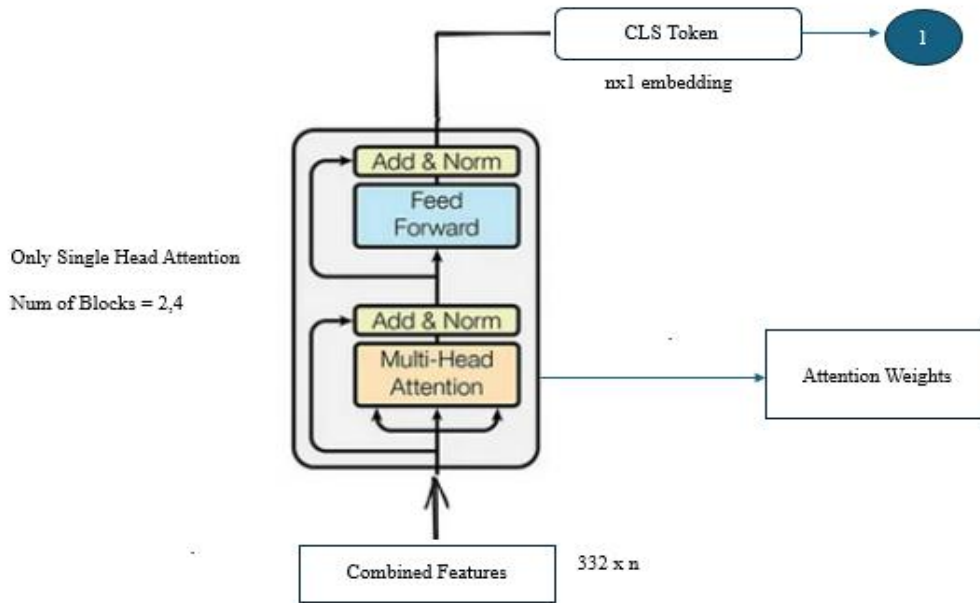


Figure 4-5 Architecture of Self Attention Stage II Training

This model replaces the baseline MLP with a Transformer-style attention encoder to better capture inter-regional dependencies among the 332 ROI embeddings derived from CNN1 and CNN2. A learnable classification token (CLS) is prepended to the concatenated  $[332 \times n]$  matrix, forming a  $[333 \times n]$  input sequence that is processed through  $L$  stacked encoder blocks (with single-head or multi-head self-attention). Each encoder block consists of a multi-head attention layer followed by a feedforward network, both wrapped with residual connections and layer normalization. Figure adapted from Vaswani et al. (2017), “Attention is All You Need

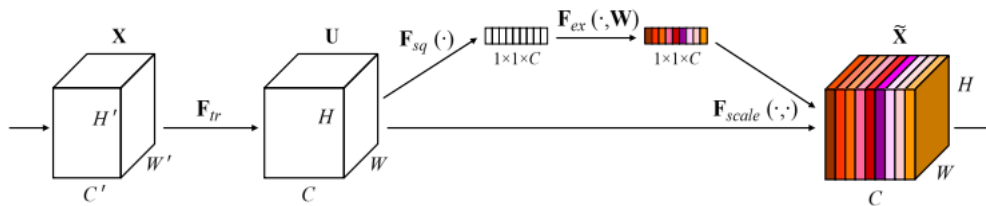
To capture higher-order interactions among ROI embeddings, the two-layer MLP regressor was replaced by a lightweight Transformer-style attention encoder topped with a classification (“CLS”) token[1] regression head. (See Figure 4-5) The pre-trained 3D-CNN encoders (CNN1 and CNN2) remain frozen; only the attention blocks and final linear layer are trained.

First, the 27 large-ROI embeddings ( $27 \times n$ ) and 305 small-ROI embeddings ( $305 \times n$ ) are concatenated into a  $332 \times n$  feature matrix. A learnable CLS token ( $1 \times n$ ) is prepended to this sequence, producing a  $(333 \times n)$  input. This sequence passes through  $L$  identical encoder blocks—

each block following the standard Transformer architecture [2]. In effect, the CLS token progressively aggregates information from all ROI embeddings via self-attention. After the final encoder, the CLS embedding ( $1 \times n$ ) is fed to a single linear layer that outputs the predicted working-memory score.

Training minimizes mean squared error using the Adam optimizer[3] with early stopping on validation loss. Only the attention encoder and regression head parameters are updated; CNN1 and CNN2 remain fixed. Learned attention weights can be inspected to identify which ROIs most strongly influence the prediction

#### 4.6 Squeeze-and-Excitation Path

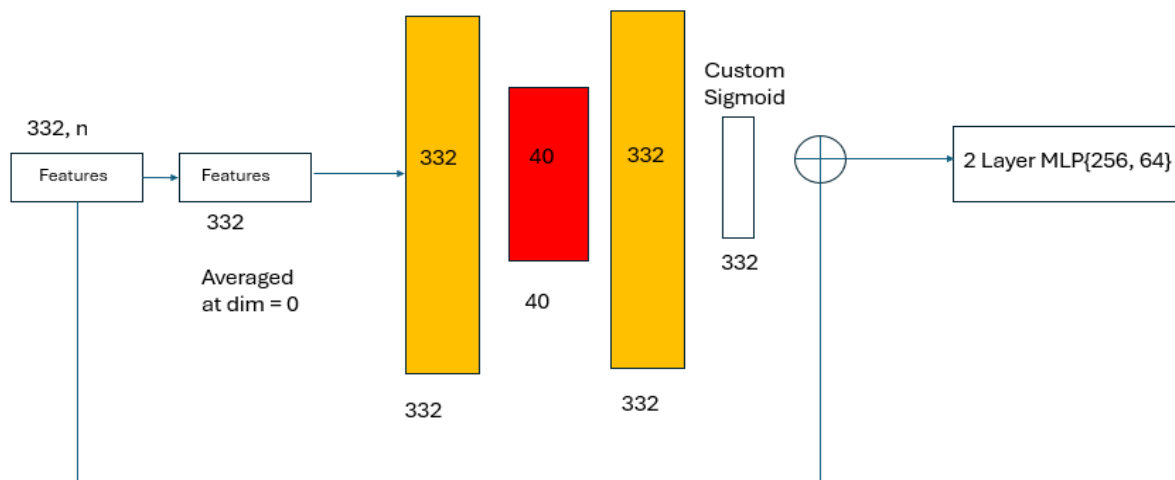


*Figure 4-6 Original Squeeze and Excitation Block*

*This module adaptively recalibrates channel-wise feature responses by first squeezing global spatial information via average pooling, then exciting channel weights through a learnable gating mechanism. The resulting weights scale the input features, enhancing the most informative channels. Figure adapted from Hu et al., “Squeeze-and-Excitation Networks,” CVPR 2018.*

The canonical SE block[4] first “squeezes” each channel’s spatial information into a single scalar via global average pooling. An “excitation” MLP then learns per-channel weights by passing these scalars through a small bottleneck (reduction and expansion) and a sigmoid nonlinearity.[See Fig 4-6] Finally, each feature-map channel is re-scaled by its learned weight, enabling the network to emphasize informative channels and suppress less useful ones. This

mechanism has been shown to improve representational power in a variety of convolutional architectures.



*Figure 4-7 Architecture of Region Squeeze and Excitation Stage II Training*  
 This figure presents a region-aware SE branch added in parallel to the CLS-token Transformer path. The  $332 \times n$  ROI feature matrix is first squeezed along the feature axis ( $\text{dim}=0$ ), producing a  $332$ -length vector summarizing each region. A bottleneck MLP ( $332 \rightarrow 40 \rightarrow 332$ ) with ReLU nonlinearity generates attention weights, which are passed through a custom sigmoid to yield region-wise gating coefficients. These weights modulate the original features via element-wise multiplication, emphasizing informative ROIs before being passed to the final regression head.

In parallel with the CLS-token Transformer path, a region-aware squeeze-and-excitation (SE) branch was added to Stage II. (See Figure 4-7) After concatenating the  $332 \times n$  feature matrix along the region axis, the per-region embeddings are squeezed by averaging over the feature dimension ( $n$ ), yielding a  $332$ -element vector that summarizes each ROI's overall activation. This vector is then passed through a small bottleneck MLP—compression to  $40$  units, ReLU, then expansion back to  $332$  units—to excite the regions. A custom sigmoid activation (with its steepness treated as a hyperparameter) transforms the MLP output into gating coefficients in

[0,1], which are multiplied element-wise with the original  $332 \times n$  features to modulate each ROI's contribution before concatenation with the CLS path and final MLP regressor.

## REFERENCES

- [1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [3] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [4] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

## 5 RESULTS I

### 5.1 Performance of GMV vs FA features

Tables 5-1 and 5-2 summarize the prediction performance of gray-matter volume (GMV) and fractional anisotropy (FA) features for four outcome measures—0-back accuracy (c0b), 2-back accuracy (c2b), CBCL attention score, and the principal-component attention score (PC1)—across three brain circuitries (full fronto-thalamo-parietal-cerebellar [FTPC], fronto-thalamo-parietal [FTP], and fronto-thalamo-cerebellar [FTC]) and three predictive models (Bayesian Ridge Regression – BRR; Support Vector Regression – SVR; Neural Network – NN). Prediction accuracy was quantified as the mean  $\pm$  SD of the correlation between observed and predicted scores over five test folds, with the top-performing model for each combination highlighted in bold

*Table 5-1 Performance of GMV Features*

<b>c0b</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	0.222 $\pm$ 0.019	0.214 $\pm$ 0.020	0.217 $\pm$ 0.019
<b>SVR</b>	<b>0.249 <math>\pm</math> 0.024</b>	<b>0.238 <math>\pm</math> 0.020</b>	<b>0.244 <math>\pm</math> 0.023</b>
<b>NN</b>	0.221 $\pm$ 0.017	0.217 $\pm$ 0.020	0.216 $\pm$ 0.020
<b>c2b</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	0.234 $\pm$ 0.009	0.225 $\pm$ 0.011	0.230 $\pm$ 0.013
<b>SVR</b>	<b>0.262 <math>\pm</math> 0.018</b>	<b>0.254 <math>\pm</math> 0.017</b>	<b>0.259 <math>\pm</math> 0.018</b>
<b>NN</b>	0.237 $\pm$ 0.011	0.226 $\pm$ 0.014	0.236 $\pm$ 0.017
<b>CBCL</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	<b>0.099 <math>\pm</math> 0.015</b>	<b>0.094 <math>\pm</math> 0.018</b>	<b>0.098 <math>\pm</math> 0.019</b>
<b>SVR</b>	0.081 $\pm$ 0.027	0.083 $\pm$ 0.030	0.083 $\pm$ 0.006
<b>NN</b>	0.084 $\pm$ 0.023	0.084 $\pm$ 0.033	0.071 $\pm$ 0.031
<b>PC1</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	0.123 $\pm$ 0.031	<b>0.119 <math>\pm</math> 0.029</b>	<b>0.117 <math>\pm</math> 0.022</b>
<b>SVR</b>	0.094 $\pm$ 0.005	0.092 $\pm$ 0.002	0.077 $\pm$ 0.016
<b>NN</b>	<b>0.128 <math>\pm</math> 0.027</b>	0.102 $\pm$ 0.037	0.099 $\pm$ 0.028

Table 5-2 Performance of FA features

<b>c0b</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	<b>0.198 ± 0.012</b>	<b>0.197 ± 0.009</b>	<b>0.185 ± 0.013</b>
<b>SVR</b>	0.191 ± 0.023	0.191 ± 0.025	0.175 ± 0.021
<b>NN</b>	0.192 ± 0.022	0.186 ± 0.008	0.180 ± 0.020
<b>c2b</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	0.214 ± 0.011	0.212 ± 0.010	<b>0.198 ± 0.014</b>
<b>SVR</b>	<b>0.219 ± 0.015</b>	<b>0.219 ± 0.015</b>	0.195 ± 0.010
<b>NN</b>	0.205 ± 0.010	0.187 ± 0.021	0.196 ± 0.012
<b>CBCL</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	<b>0.091 ± 0.025</b>	<b>0.083 ± 0.027</b>	<b>0.072 ± 0.021</b>
<b>SVR</b>	0.058 ± 0.021	0.055 ± 0.023	0.043 ± 0.028
<b>NN</b>	0.074 ± 0.037	0.059 ± 0.041	0.048 ± 0.031
<b>PC1</b>	<b>FTPC</b>	<b>FTP</b>	<b>FTC</b>
<b>BRR</b>	<b>0.119 ± 0.031</b>	<b>0.113 ± 0.031</b>	0.095 ± 0.035
<b>SVR</b>	0.094 ± 0.008	0.055 ± 0.023	<b>0.096 ± 0.010</b>
<b>NN</b>	0.115 ± 0.030	0.099 ± 0.042	0.058 ± 0.021

Across all models, GMV features yielded their strongest predictions for cognitive measures:

- c0b: SVR (FTPC =  $0.249 \pm 0.024$ ; FTP =  $0.238 \pm 0.020$ ; FTC =  $0.244 \pm 0.023$ ).
- c2b: SVR (FTPC =  $0.262 \pm 0.018$ ; FTP =  $0.254 \pm 0.017$ ; FTC =  $0.259 \pm 0.018$ ).

BRR and NN showed similar but slightly lower correlations (e.g. c2b, BRR FTPC =  $0.234 \pm 0.009$ ; NN FTPC =  $0.237 \pm 0.011$ ). Importantly, FTPC features marginally outperformed both FTP and FTC across models for both c0b and c2b. Prediction accuracies using FA features were lower than for GMV but followed the same hierarchy of circuitries:

- c0b: SVR FTPC =  $0.191 \pm 0.023$ ; FTP =  $0.191 \pm 0.025$ ; FTC =  $0.175 \pm 0.021$ .
- c2b: SVR FTPC =  $0.219 \pm 0.015$ ; FTP =  $0.219 \pm 0.015$ ; FTC =  $0.195 \pm 0.010$ .

Across models, FTPC again led, with FTP close behind and FTC slightly lower.

Behavioral predictions remained minimal for all groups and models (CBCL ; PC1).

## 5.2 Meta Analysis

*Table 5-3 Meta-Analysis of feature group performances FTPC vs FTP vs FTC*

Target	Comparison	GMV		FA	
		Z score	P value	Z score	P value
c0b	FTPC vs FTP	4.5	<0.001	1.23	0.219
	FTPC vs FTC	3.89	<0.001	10.78	<0.001
	FTC vs FTP	0.74	0.459	-12.12	<0.001
c2b	FTPC vs FTP	11	<0.001	1.97	0.048
	FTPC vs FTC	6.45	<0.001	11.86	<0.001
	FTC vs FTP	2.24	<b>0.029</b>	-12.14	<b>&lt;0.001</b>

We further conducted meta-analysis to compare FTC's prediction accuracies with those of FTP and FTPC across all predictive models and thirty testing/validation datasets, such that one meta-analysis was based on 180 observations. (See table 5-3) The inverse variance algorithm implemented in Python Meta tool was used for the analysis with mean difference as the effect measure.

We conducted only one meta-analysis for each null hypothesis (e.g., no significant difference in prediction accuracies between the FTC and FTP circuitry). Consequently, no further correction for multiple tests was applied to the meta-analysis results.

The results indicated that the full FTPC circuitry overall outperformed FTP and FTC, with significantly higher prediction accuracies noted for both c0b and c2b using either GMV or FA features as predictors, except when compared with FTP in predicting c0b using FA features. It was also noted that the FTC circuitry consistently showed significantly higher accuracies than FTP (highlighted in bold in Table 5-3) in predicting c2b regardless of GMV or FA features

### 5.3 ROI Interpretation and Discussion

Based on the kneeling point of feature importance, we examined the top 30 important ROIs for each model. Based on the common ROIs of three predictive models, we further compared the FTPC circuitry with FTC in terms of the top important ROIs. The corresponding spatial maps are shown in Figure 5.1-1d to 1g.

Figure 5-1 compares the brain regions most predictive of vigilance (0-back accuracy, c0b) and working memory (2-back accuracy, c2b) across three circuits: FTPC (frontal, thalamus, parietal, cerebellum), FTP (frontal, thalamus, parietal), and FTC (frontal, thalamus, cerebellum). For vigilance, a core set of bilateral prefrontal ROIs—middle, inferior, and superior frontal gyri plus medial and rectal gyri—emerges across all models. Augmenting this with parietal lobules (inferior and superior) in FTPC refines prediction, highlighting a fronto-parietal scaffold. In contrast, FTC-only models emphasize an anterior cingulate–cerebellar motif, featuring the cingulate gyrus, orbital frontal gyrus, declive, inferior semi-lunar lobule, and thalamus, underscoring cerebellar contributions to sustained attention.

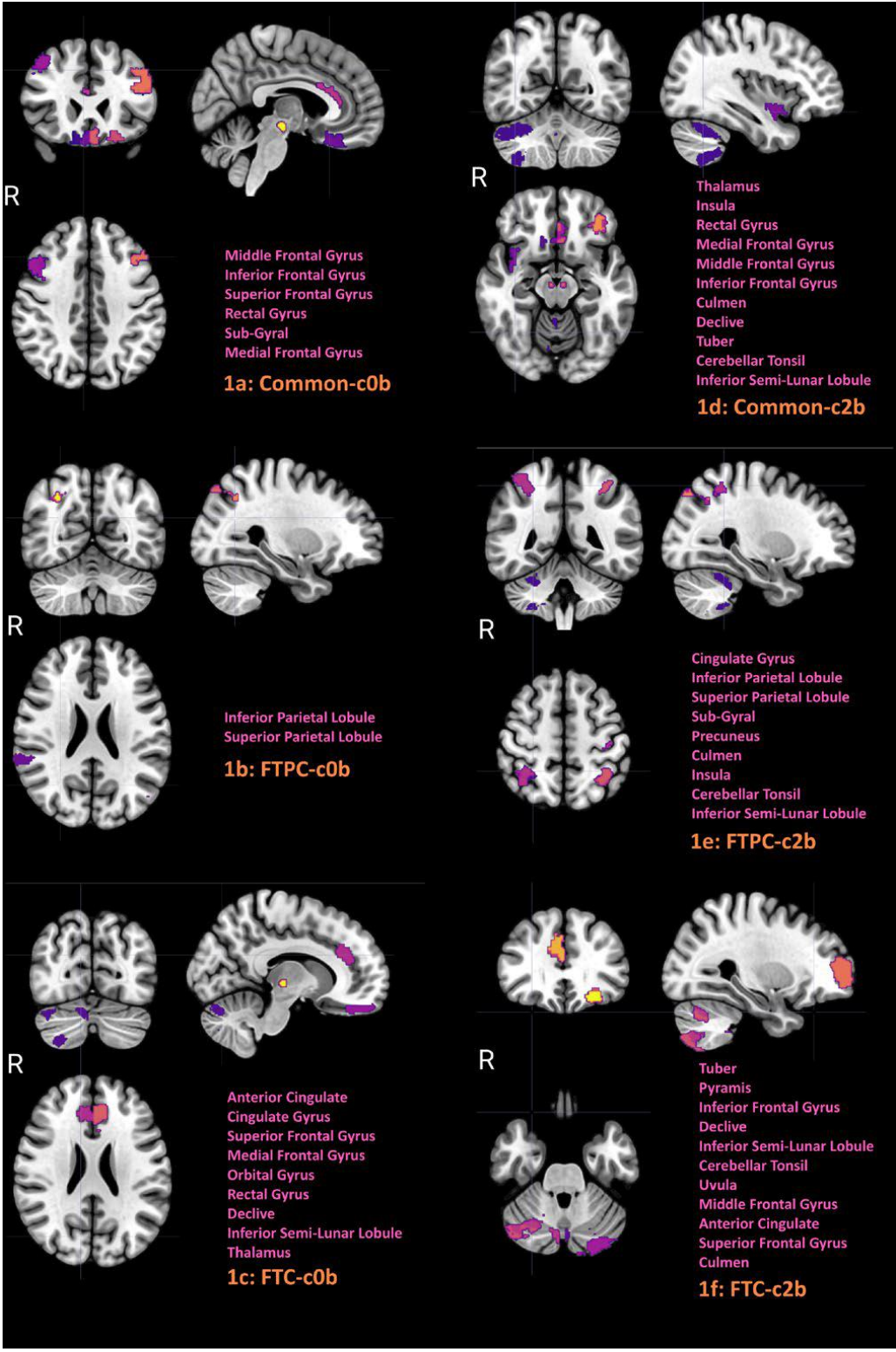


Figure 5-1 Top Regions from FTC and FTPC  
This figure depicts key brain regions (ROIs) that were most predictive of 0-back (c0b; attention) and 2-back (c2b; working memory) performance, separated by circuit type: FTPC (Fronto-

*Thalamo-Parietal-Cerebellar), FTC (Fronto-Thalamo-Cerebellar), and regions common to both. Panels 1a and 1d highlight ROIs identified by all models across both circuitries for c0b and c2b, respectively—showcasing consistent involvement of frontal regions, thalamus, and select cerebellar lobules. Panels 1b and 1e show FTPC-specific ROIs for c0b and c2b, including parietal lobules, cingulate cortex, and precuneus. Panels 1c and 1f present FTC-specific ROIs, enriched in medial/orbital frontal cortex, rectal gyrus, and cerebellar regions such as declive, uvula, and inferior semi-lunar lobule.*

For memory, shared ROIs again include frontal gyri, insula, thalamus, and cerebellar nodules (culmen, declive, tuber, tonsil, inferior semi-lunar lobule). Extending to FTPC recruits additional parietal and midline regions—cingulate gyrus, precuneus, inferior/superior parietal lobules—suggesting that parietal-cerebellar interactions support maintenance processes. The FTC-only model isolates cerebellar (tuber, pyramis, uvula, declive) and prefrontal (middle/superior frontal, anterior cingulate) ROIs with insula, demonstrating that frontal-cerebellar loops alone can predict working-memory accuracy. Overall, vigilance relies on a distributed frontal core with modality-specific cerebellar or parietal enhancements, whereas working memory engages an integrated frontoparietal-cerebellar network whose parietal nodes enhance global performance and whose cerebellar nodes uniquely fine-tune maintenance.

## 6 RESULTS II

### 6.1 Training Results

Table 6.1 reports cross-validated performance when concatenating the first N features (25, 50, 75, 100) from the CNN encoders. Validation  $R^2$  remains stable around 0.08–0.083 and Pearson  $r \approx 0.29$ , indicating consistent fit across feature set sizes. On the held-out test set, the 25-feature model achieves the highest  $R^2$  (0.073) and  $r$  (0.27), with performance declining as more features are added (e.g.,  $R^2=0.051$ ,  $r=0.24$  at  $N=75$ ). This suggests that a more compact embedding may generalize better than larger, potentially noisier representations

*Table 6-1 Baseline Concatentaion Training Performance*

Features	Val R2	Val Corr	Test R2	Test Corr
25	0.0831	0.29	0.073	0.27
50	0.0828	0.29	0.068	0.26
75	0.0817	0.29	0.051	0.24
100	0.0785	0.28	0.059	0.25

Introducing region-level squeeze-and-excitation (Table 6.2) further boosts predictive accuracy. An  $8\times$  compression yields validation  $R^2=0.093$  ( $r=0.308$ ) and test  $R^2=0.079$  ( $r=0.28$ ), outperforming the baseline across all metrics. In contrast, a more aggressive  $16\times$  squeeze

underfits (test  $R^2=0.056$ ,  $r=0.23$ ), indicating that retaining sufficient channel capacity is critical for capturing region-specific signal.

*Table 6-2 Region Squeeze and Excitation Performance*

Features	Val R2	Val Corr	Test R2	Test Corr
8x Squeeze	0.0931	0.3081	0.079	0.28
16x Squeeze	0.0618	0.2478	0.056	0.23

Table 6.3 compares transformer encoder depths. Two layers of self-attention improve on four layers at test time ( $R^2=0.052$ ,  $r=0.23$  vs.  $R^2=0.042$ ,  $r=0.20$ ), while validation metrics mirror this trend. This suggests that a shallow attention stack strikes the best balance between modeling higher-order feature interactions and avoiding overfitting.

*Table 6-3 Self Attention Encoder Performance*

Num of Blocks	Val R2	Val Corr	Test R2	Test Corr
4 Blocks	0.0640	0.26	0.042	0.20
2 Blocks	0.0741	0.28	0.052	0.23

Across all experiments, the 8× squeeze-and-excitation model emerges as the strongest, achieving the highest generalization to unseen data.

## 6.2 Saliency Maps and Significant Volumes

Using normalized saliency scores from the top-performing 8× squeeze-and-excitation model, the thirty most important ROI features were identified via histogram cutoff. Each ROI's binary saliency mask was transformed into Talairach space and its left- and right-hemisphere volumes (in cm<sup>3</sup>) quantified. To ensure robustness, only regions whose combined bilateral volume exceeded 1 cm<sup>3</sup> were retained (except the thalamus, given its small size). The final selection spanned frontal, parietal, and cerebellar territories (Table 6-4), revealing a distributed network in which both frontoparietal executive nodes and multiple cerebellar lobules jointly support working-memory prediction.

*Table 6-4 Anatomical regions and bilateral saliency-mask volumes (left / right hemisphere, in cc)*

Area	volume (cc)	Region
Superior Frontal Gyrus	0.8/7.7	Frontal
Middle Frontal Gyrus	0.0/3.4	Frontal
Thalamus	0.0/0.2	thalamus
Culmen	2.2/0.0	Cerebellum
Declive	5.8/1.1	Cerebellum
Uvula	2.7/1.2	Cerebellum
Pyramis	3.5/2.0	Cerebellum
Tuber	4.6/2.6	Cerebellum
Cerebellar Tonsil	2.6/1.1	Cerebellum

Inferior Semi-Lunar Lobule	2.8/1.1	Cerebellum
Medial Frontal Gyrus	1.5/3.3	Frontal
Anterior Cingulate	0.1/1.1	Frontal
Inferior Frontal Gyrus	0.0/2.9	Frontal
Postcentral Gyrus	0.4/1.7	Parietal

### 6.3 Regions and Literature Discussion

The regions highlighted align closely with core sensory and attentional hubs known to support working memory and executive function. In the following subsections, each node's role is examined considering established neurophysiological and cognitive findings.

#### 6.3.1 Parietal and Thalamus

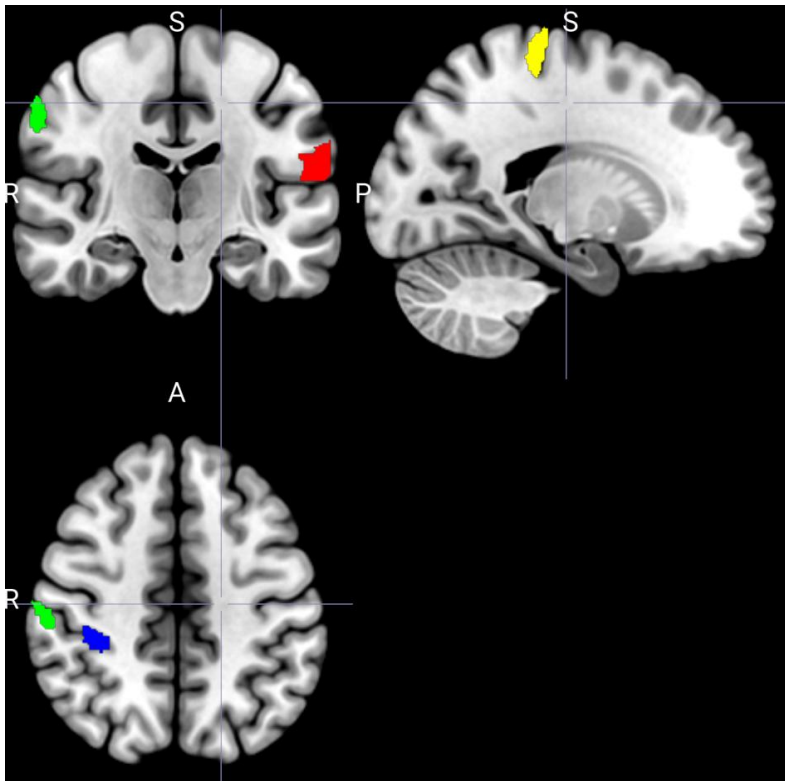


Figure 6-1 Parietal Regions: Postcentral Gyrus and Thalamus

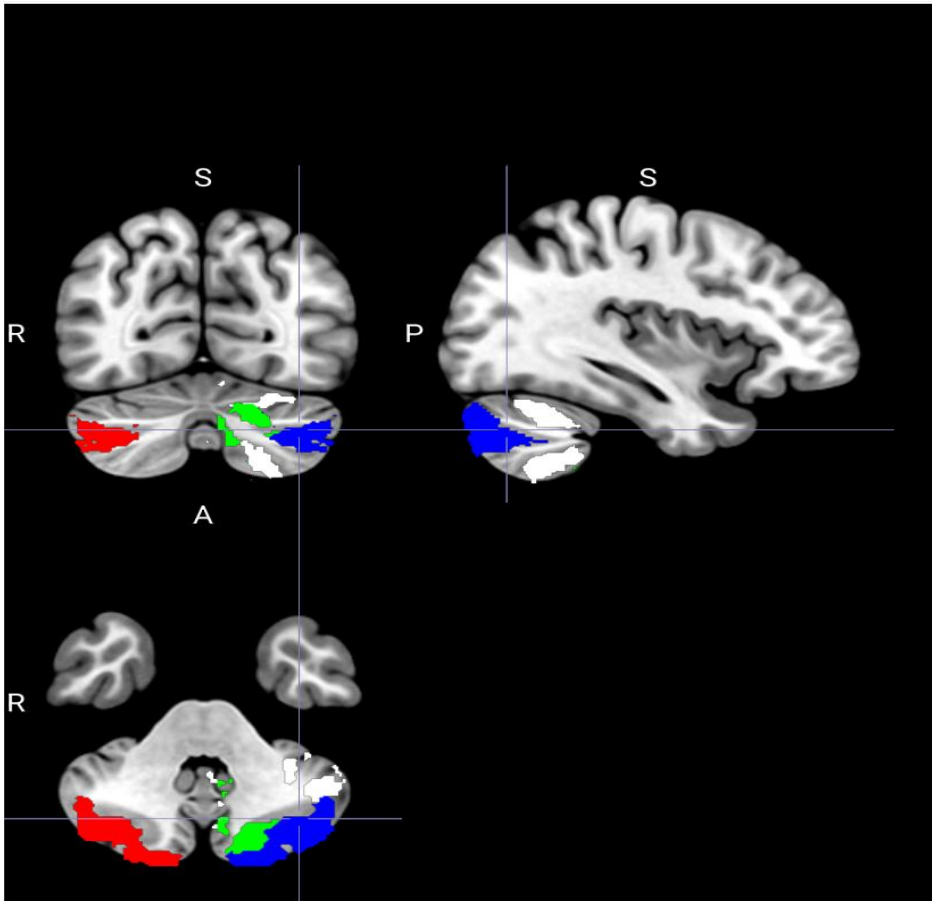
The postcentral gyrus—the primary somatosensory cortex—processes tactile and proprioceptive inputs, creating a somatotopic map essential for discerning touch, pressure, and limb position. In working memory tasks, its sensory representations help maintain relevant sensorimotor context and filter incoming cues.

The thalamus acts as a gateway for nearly all sensory signals to the cortex and orchestrates cortico-subcortical loops. By filtering and synchronizing inputs, it supports attentional selection and network timing. During executive tasks, thalamic gating ensures that only task-relevant information reaches prefrontal regions, while thalamo-cortical rhythms coordinate the distributed activity underpinning working memory and cognitive control[1].

### **6.3.2 *Cerebellar Lobules***

The cerebellar vermis and hemispheric lobules highlighted by the squeeze-and-excitation model—including the uvula, tonsil, tuber, pyramis, declive, and inferior semi-lunar lobule—are increasingly recognized for their roles beyond basic motor coordination. The uvula and tonsil, located in the vestibulocerebellum, contribute to postural control and the integration of vestibular and proprioceptive signals, which may underpin the maintenance of spatial attention and sensorimotor context in working-memory tasks[2]. The tuber and pyramis, part of the spinocerebellar, receive somatosensory input and are implicated in fine motor planning and error correction, processes that parallel the on-line updating of information during cognitive load. Finally, the declive and inferior semi-lunar lobule within the cerebrocerebellar relate to prefrontal and parietal cortices and have been shown to modulate executive functions, particularly

the timing and sequencing of mental operations[2]. Collectively, these lobules support a cerebellar contribution to cognition by coordinating sensorimotor precision, attentional stability, and the temporal organization of working-memory processes.

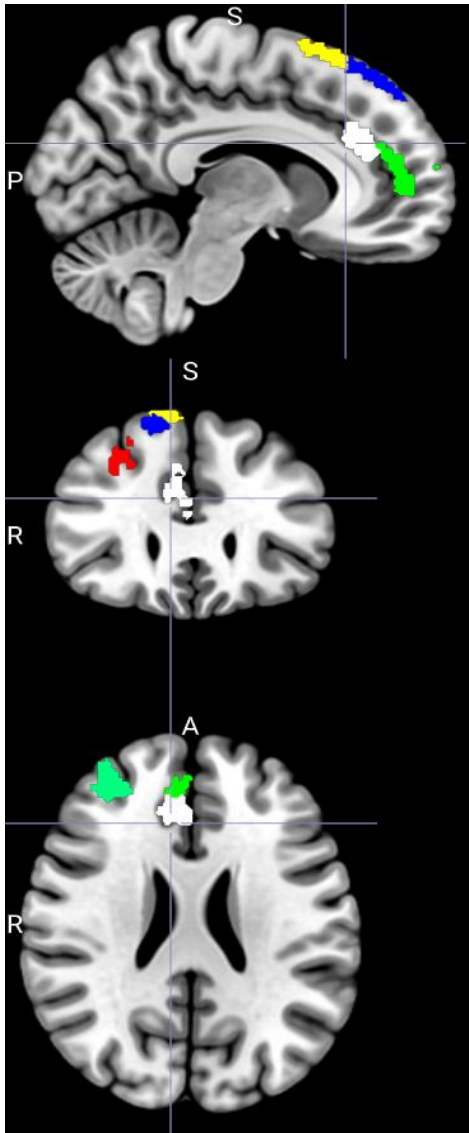


*Figure 6-2 Top Cerebellar Lobules; Uvula, Cerebellar Tonsil, Tuber, Pyramis, Declive, Inferior Semi-Lunar Lobule*

### **6.3.3 Prefrontal Coretx**

(Figure 6-2 )The superior frontal gyrus (yellow) contributes to high-level executive processes such as working-memory maintenance and prospective planning, integrating information across distributed cortical circuits[3]. The middle frontal gyrus (blue) supports the manipulation and monitoring of information in working memory and

underlies cognitive flexibility during task switching. Medial frontal regions (green), including portions of the medial frontal gyrus, are central to self-referential evaluation and the allocation of attentional resources, helping to sustain goal-directed behavior. The anterior cingulate cortex (white) exerts conflict monitoring and error-detection functions, dynamically adjusting control signals to optimize performance under load. Finally, the inferior frontal gyrus (red) is critical for inhibitory control and the selection of task-relevant representations, gating irrelevant stimuli to prevent interference. Together, these frontal top nodes form a coordinated executive network that orchestrates the encoding, maintenance, and manipulation of working-memory contents.[3]



*Figure 6-3 Frontal Top Regions: Yellow: Superior Frontal Gyrus, Blue: Middle Frontal Gyrus, Green(both): Medial Frontal Gyrus, White: Anterior Cingulate, Red: Inferior Frontal Gyrus*

## REFERENCES

- [1] Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 357(1428), 1695–1708.
- [2] Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *NeuroImage*, 44(2), 489–501.
- [3] Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

## 7 CONCLUSION

Early adolescence working memory hinges on a finely tuned circuit that bridges the cerebellum, thalamus, parietal cortex, and frontal lobes. In Project I, simple anatomical measures—gray-matter volume in cerebellar lobules (declive, pyramis), thalamic nuclei, the postcentral gyrus, and multiple frontal gyri—consistently tracked performance on both vigilance (0-back) and working-memory (2-back) tasks ( $r \approx .25-.26$ ). Notably, combining all four nodes into a fronto-thalamo-parietal-cerebellar network outperformed models that omitted either the parietal or cerebellar components, underscoring the importance of each region in the executive scaffold.

Project II then wove a richer tapestry. Rather than averaging entire regions, three-dimensional T1-patches were carved out around each anatomical landmark, clustered by size, and fed into bespoke 3D-CNN encoders. Their latent outputs were merged via a squeeze-and-excitation mechanism that adaptively weighted each region's contribution. This approach lifted predictive power by about 25 percent over the classical volume-only models ( $R^2 = .079$ ,  $r = .28$ ), and its saliency maps lit up the very same cerebellar (declive, uvula, tuber, pyramis, inferior semi-lunar lobule, tonsil), thalamic, parietal, and frontal (superior, middle, medial frontal gyri; anterior cingulate) subregions flagged in Project I.

Both studies point to the same set of brain regions working together as a team to support working memory in young teens.

Putting these findings together suggests there's a clear pathway:

1. Cerebellum sends finely tuned timing and sensor information up to
2. Thalamus, which gates and coordinates those signals before passing them on to
3. Parietal cortex (postcentral gyrus), where sensory details are integrated, and finally to

4. Frontal cortex (superior, middle, medial, anterior cingulate, and inferior frontal gyri), where the brain holds, updates, and manipulates information in working memory.

In other words, a loop from cerebellum → thalamus → parietal → frontal is a possibly a backbone in working memory tasks for adolescents.

# APPENDICES

## Appendix A

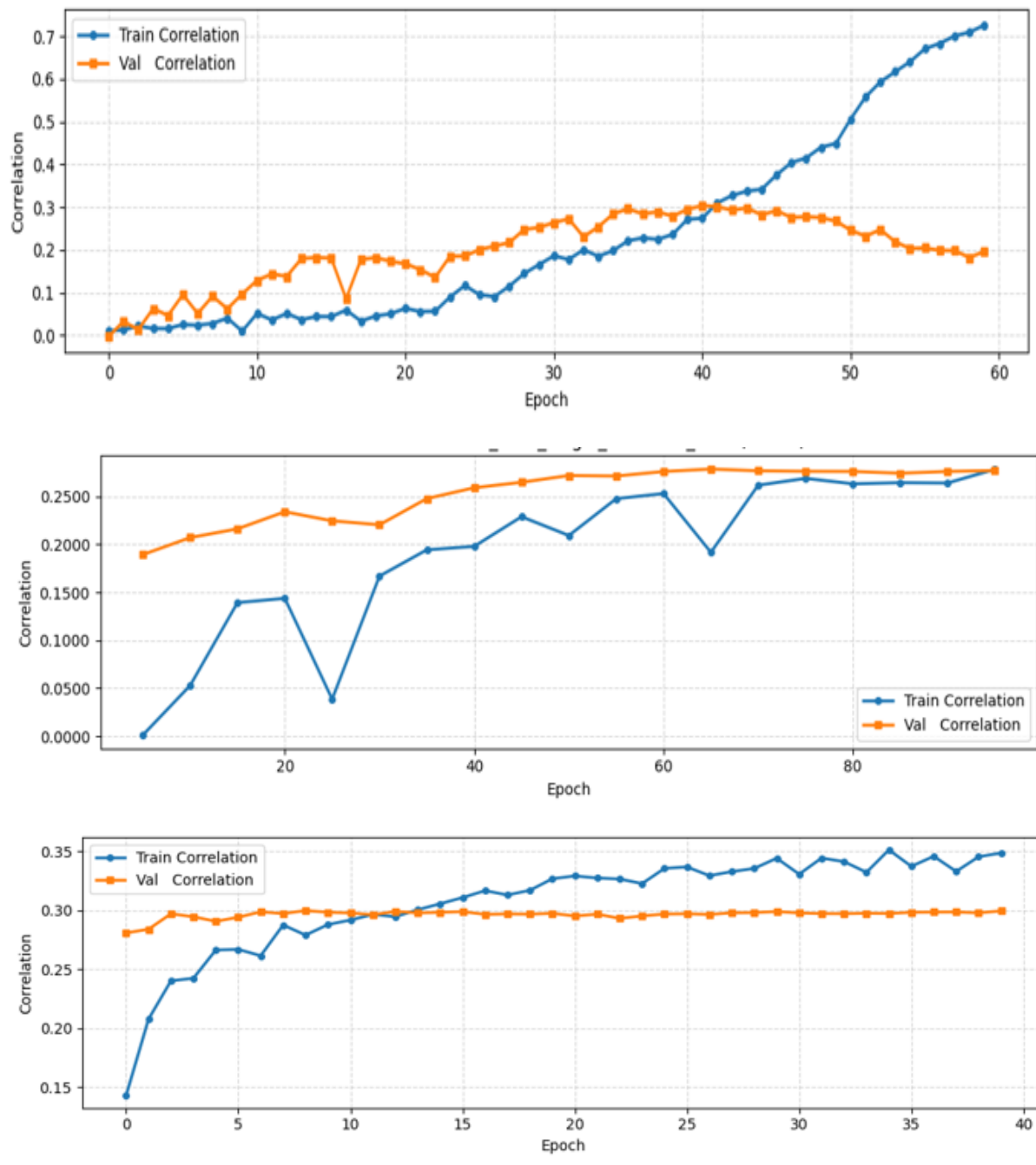


Figure 7-1 Training Curves of Baseline, Self Attention and Squeeze and Excitation Modules from up to bottom

Module	Operation	Parameters
<b>CNN1</b>		
Conv3D (k=3)	conv3d	in=1, out=32, kernel=3, stride=1
ReLU	activation	–
MaxPool3D (k=2)	pooling	kernel =2
Dropout3D	regularization	p=0.5 (default)
Conv3D (k=3)	conv3d	in=32, out=64, kernel =3, stride=1
Normalization + ReLU	batch/group/layer norm + activation	num_params =64
MaxPool3D (k=2)	pooling	kernel=2
Flatten + Linear	dense	in=4096, out=n
<b>CNN2</b>		
Conv3D (k=3)	conv3d	in=1, out=32, kernel=3, stride=1
ReLU	activation	–
MaxPool3D (k=2)	pooling	kernel=2
Dropout3D	regularization	p=0.5 (default)
Conv3D (k=3)	conv3d	in=32, out=64, kernel=3, stride=1
Normalization + ReLU	batch/group/layer norm + activation	num_params=64
MaxPool3D (k=3)	pooling	kernel=3
Flatten + Linear	dense	in=8000, out=n
<b>Regressor</b>		
Linear	dense	in=332 · n, out=256
ReLU + Dropout	activation + regularization	p=0.5 (default)
Linear	dense	in=256, out=64
ReLU + Dropout	activation + regularization	p=0.5 (default)
Linear	dense	in=64, out=1

*Table 7-1 Parameters of CNN1, CNN2, MLP Regressor*