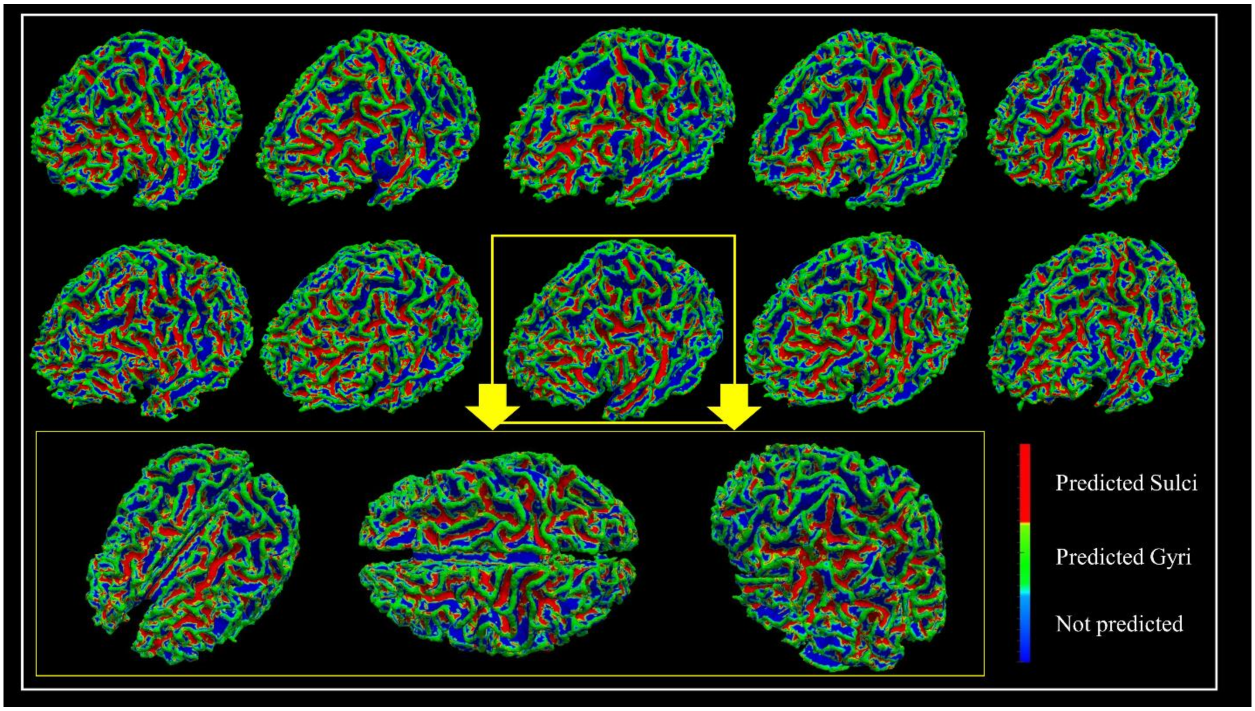
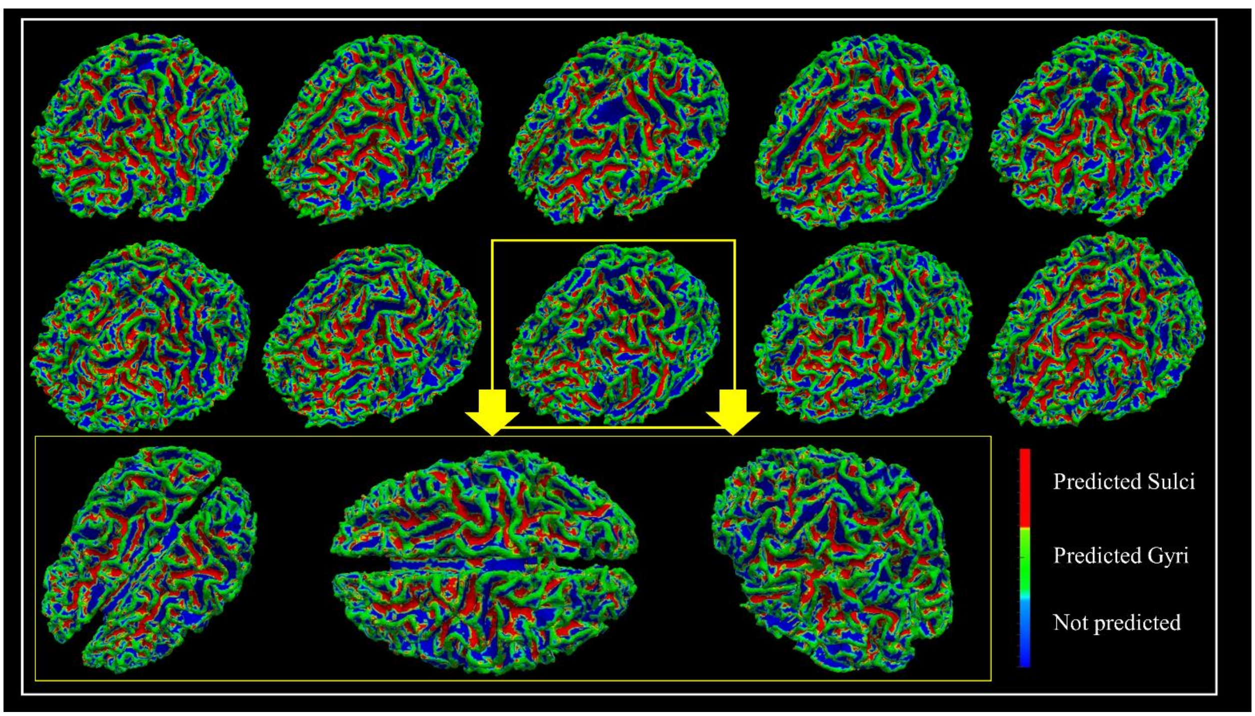
Supplemental Materials

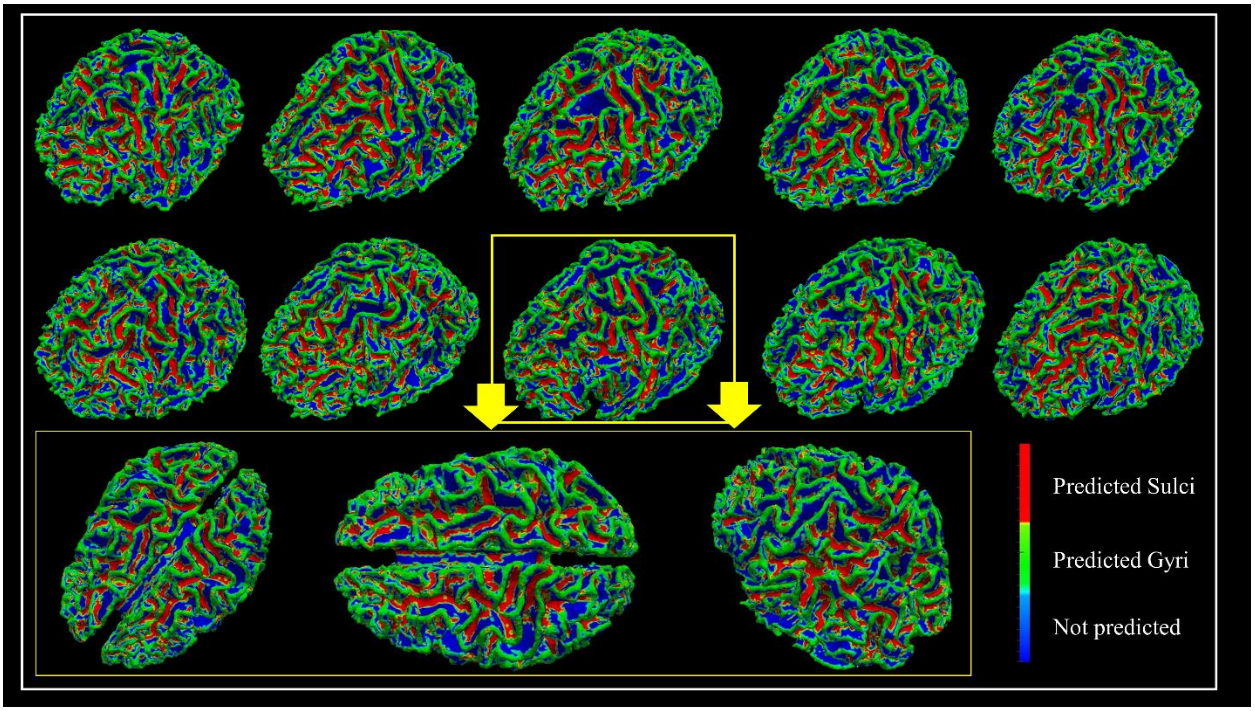
Supplemental material includes 22 figures and 1 table. They are shown as follows.



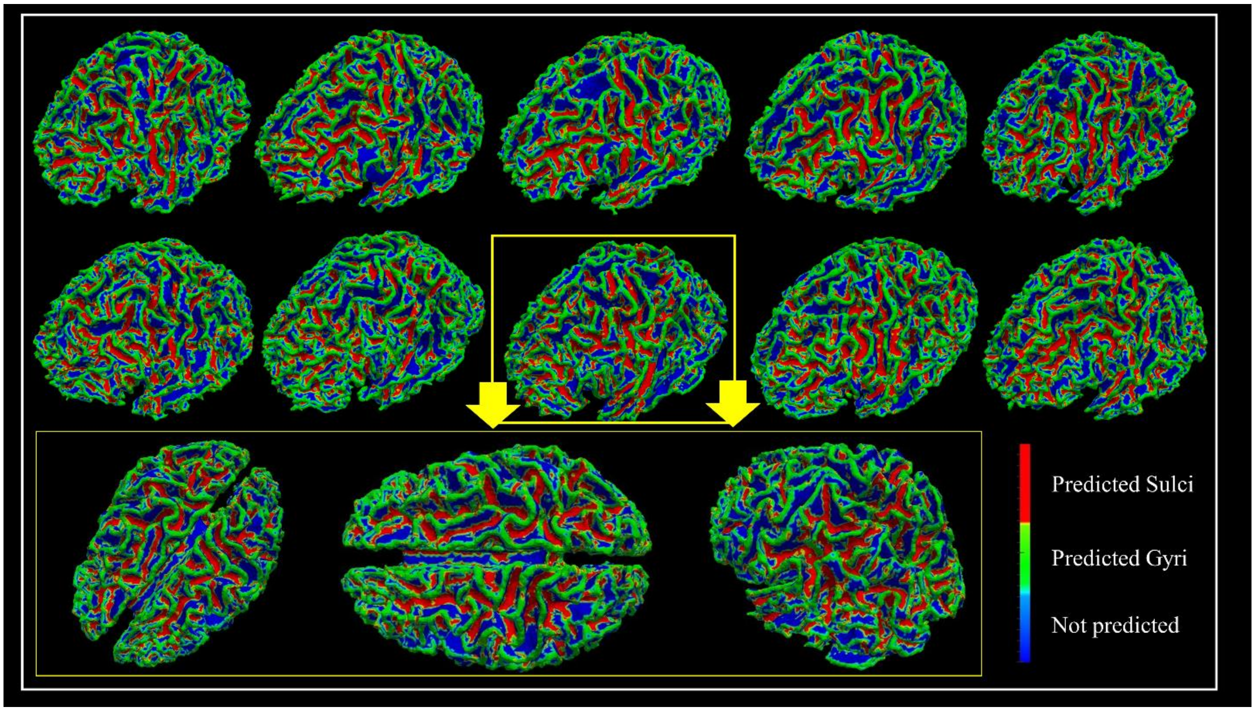
Supplemental Figure 1. The classification performance of 10 subjects in task GAMBLING.



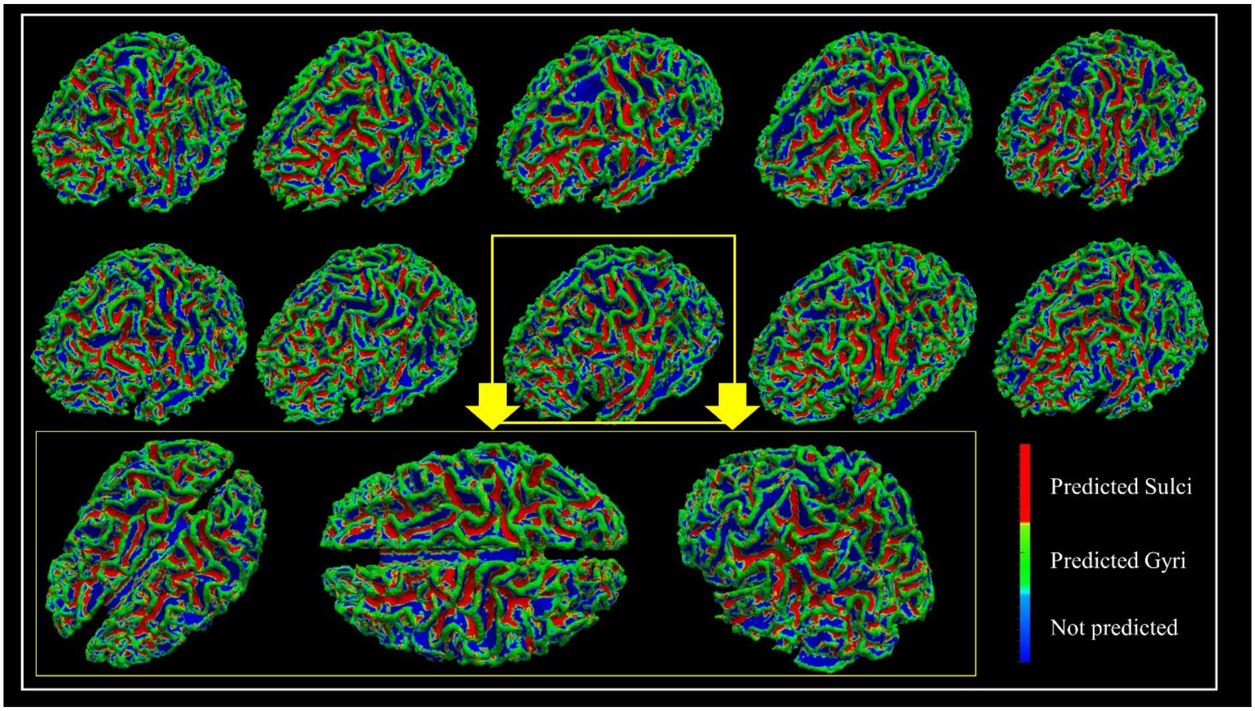
Supplemental Figure 2. The classification performance of 10 subjects in task LANGUAGE.



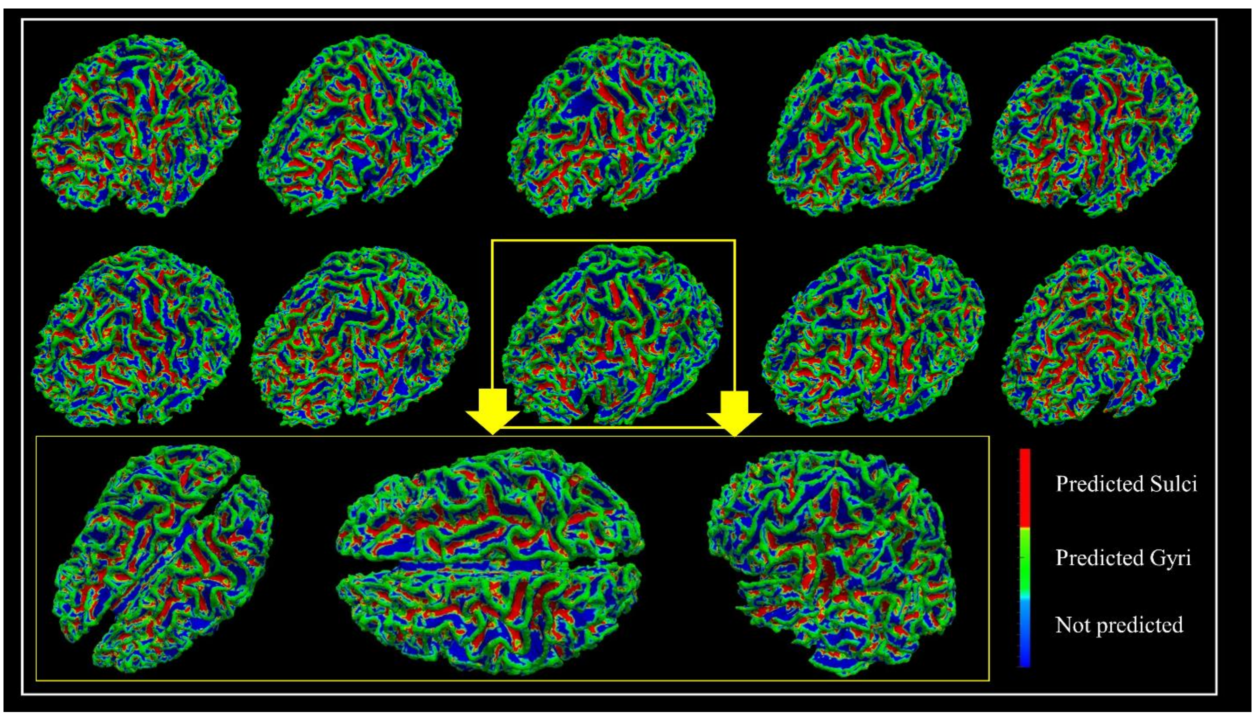
Supplemental Figure 3. The classification performance of 10 subjects in task MOTOR.



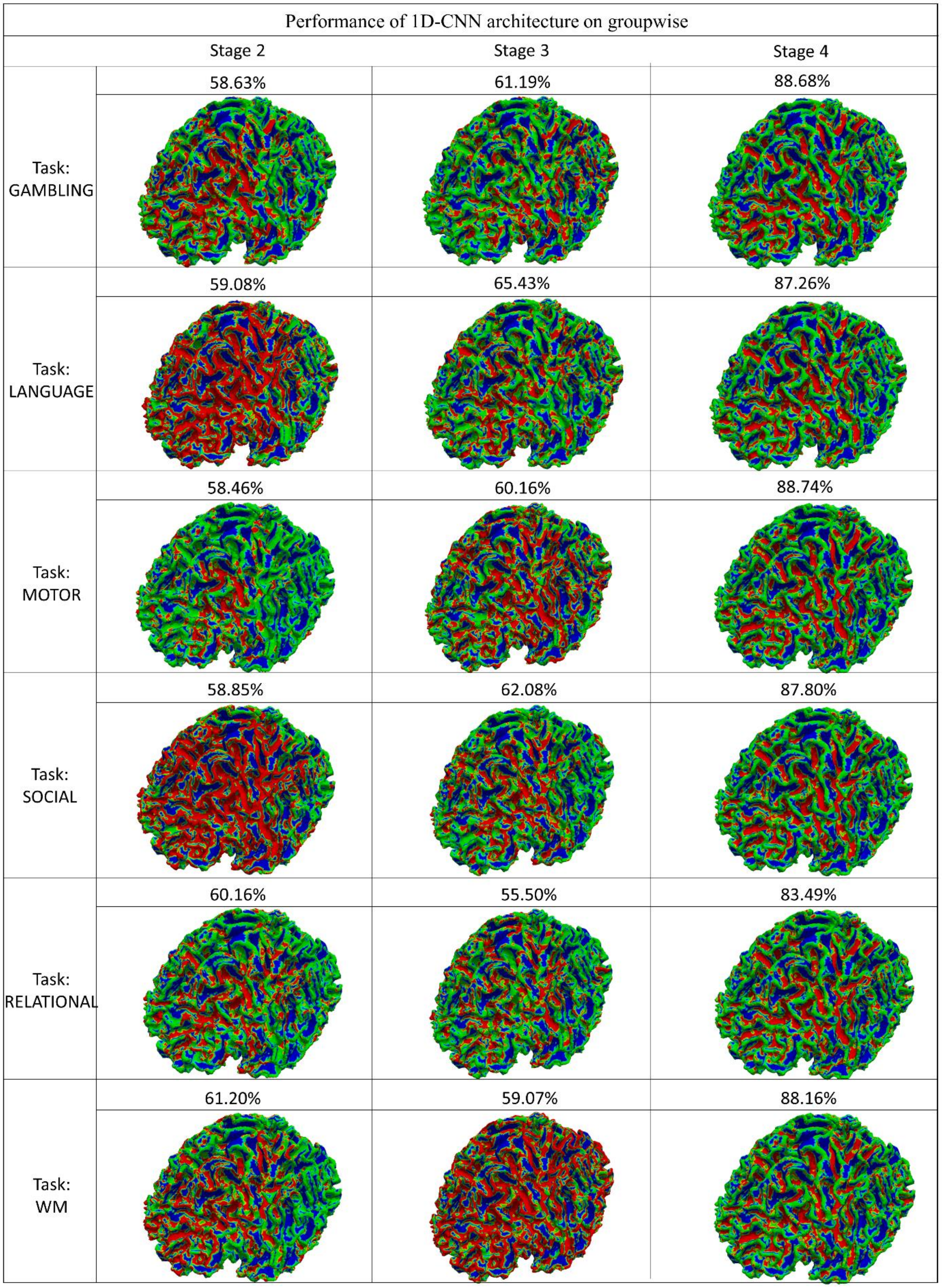
Supplemental Figure 4. The classification performance of 10 subjects in task RELATIONAL.



Supplemental Figure 5. The classification performance of 10 subjects in task SOCIAL.



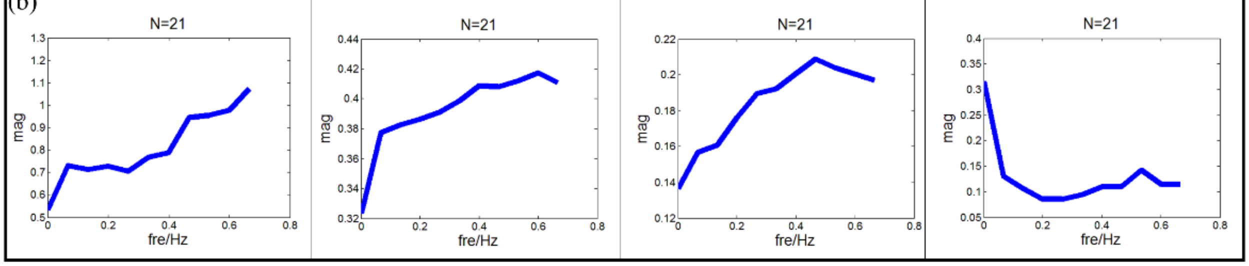
Supplemental Figure 6. The classification performance of 10 subjects in task WM.



Supplemental Figure 7. The classification performance of 4 convolutional layers in six tasks. Each voxel is given a predicted label. Spatial map of 2 different directions are used to show the classification performance of each layer. Classification accuracies are recorded, respectively.

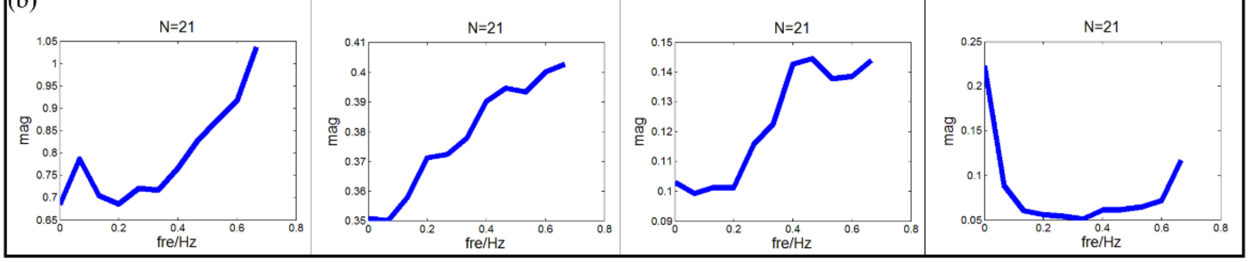
Green voxels are predicted as gyri, red voxels are predicted as sulci and blues ones are voxels that not used.

Performance of 1D-CNN architecture (Task GAMBLING)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	55.22%	58.63%	61.19%	88.68%
(b)				



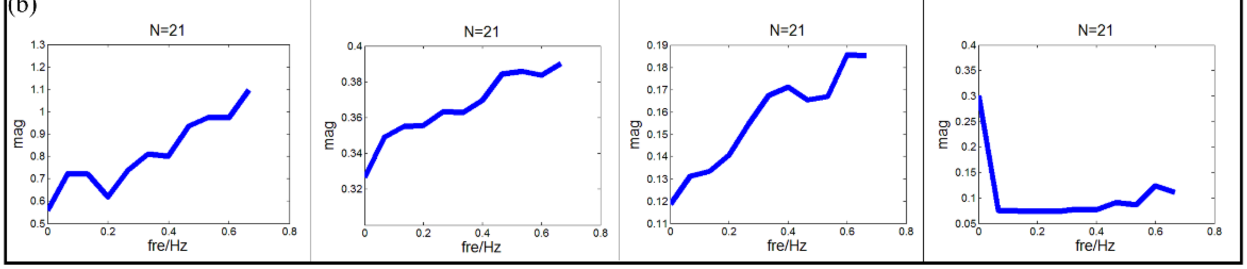
Supplemental Figure 8. Feature analysis of task GAMBLING. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task LANGUAGE)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	52.71%	59.08%	65.43%	87.26%
(b)				



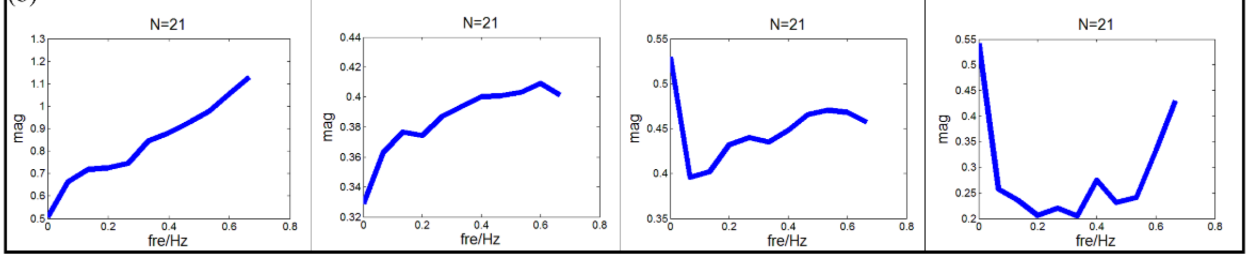
Supplemental Figure 9. Feature analysis of task LANGUAGE. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task MOTOR)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	54.25%	58.46%	60.16%	88.74%
(b)				



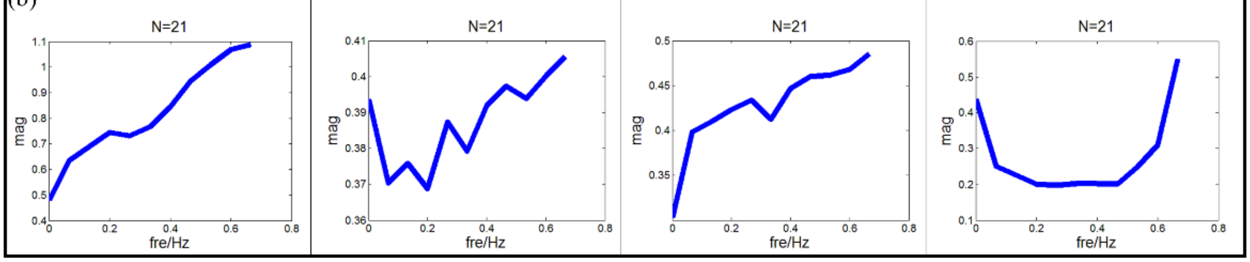
Supplemental Figure 10. Feature analysis of task MOTOR. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task RELATIONAL)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	55.79%	60.16%	55.50%	83.49%
(b)				



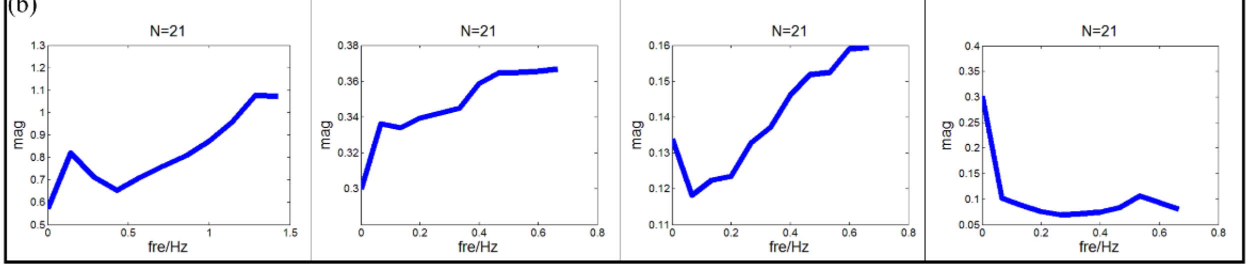
Supplemental Figure 11. Feature analysis of task RELATIONAL. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task SOCIAL)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	55.32%	58.85%	62.08%	87.80%
(b)				

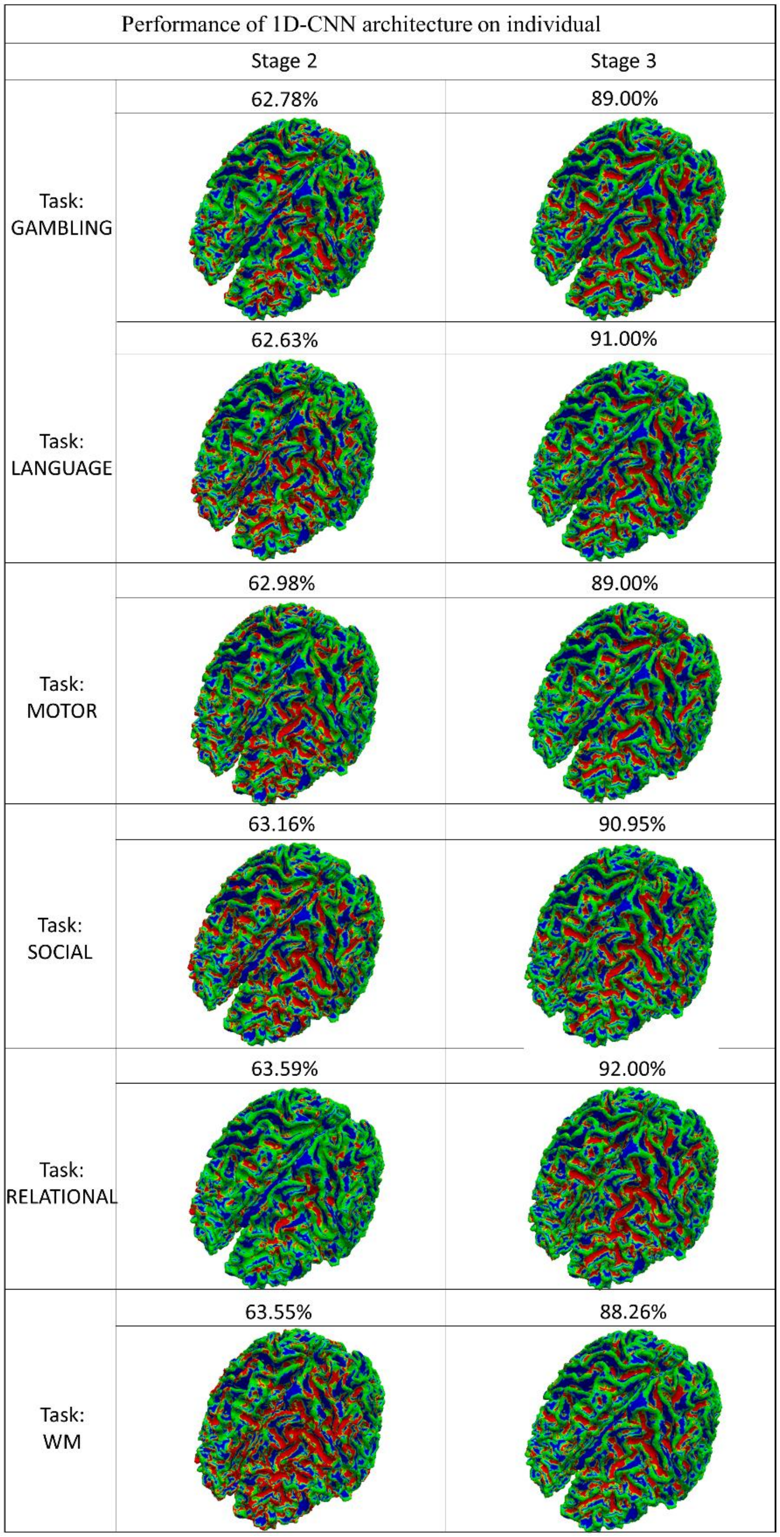


Supplemental Figure 12. Feature analysis of task SOCIAL. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task WM)				
	Layer 1	Layer 2	Layer 3	Layer 4
(a)	55.40%	61.20%	59.07%	88.16%
(b)				

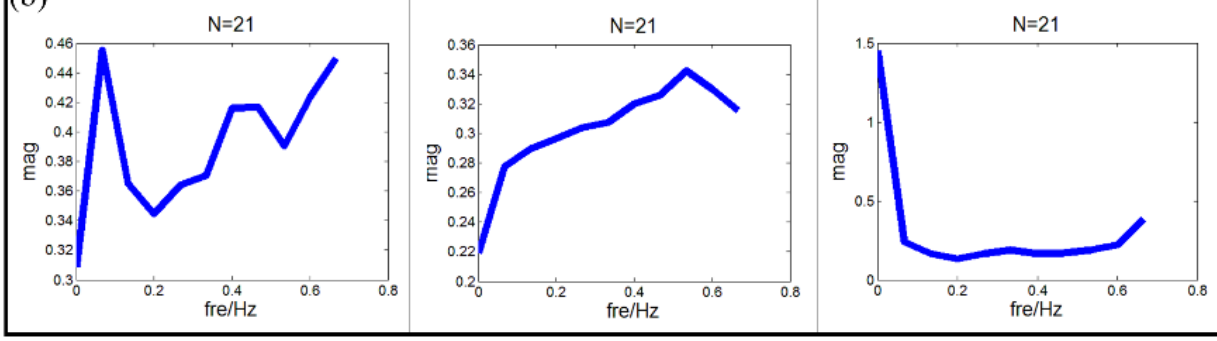


Supplemental Figure 13. Feature analysis of task WM. (a) classification performance. (b) the distribution of frequency for features in average level.



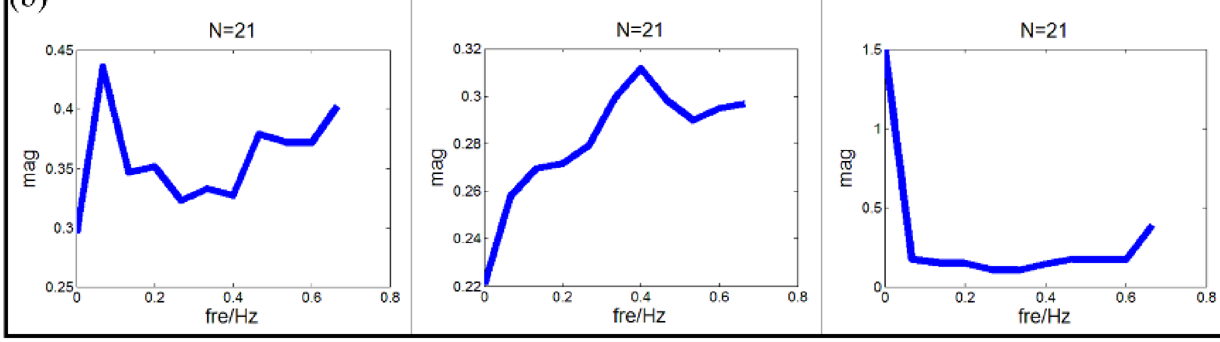
Supplemental Figure 14. The classification performance of 3 convolutional layers in six tasks. Last two layers are presented. Each voxel will be given a predicted label. Spatial map of 2 different directions are used to show the classification performance of each layer. Classification accuracies are recorded, respectively. Green voxels are predicted as gyri, red voxels are predicted as sulci and blues ones are voxels that not used.

Performance of 1D-CNN architecture (Task GAMBLING)			
	Layer 1	Layer 2	Layer 3
(a)	55.45%	62.78%	89.00%
(b)			



Supplemental Figure 15. Feature analysis of task GAMBLING. (a) classification performance. (b) the distribution of frequency for features in average level.

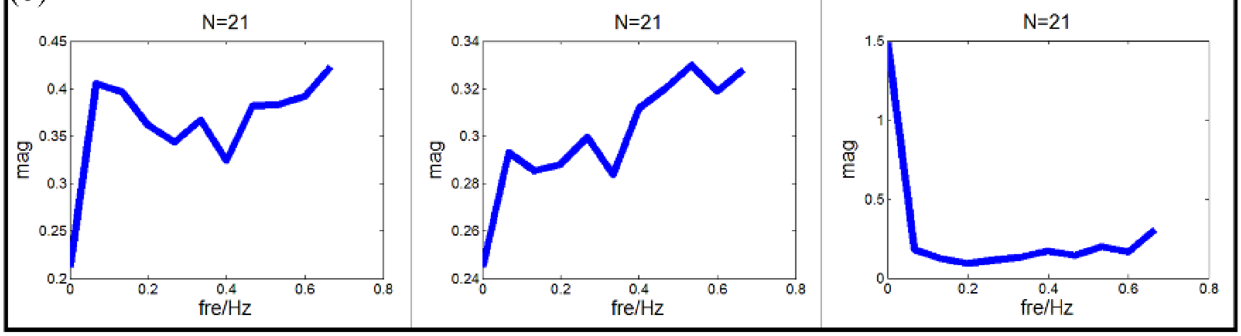
Performance of 1D-CNN architecture (Task LANGUAGE)			
	Layer 1	Layer 2	Layer 3
(a)	52.98%	62.63%	91.00%
(b)			



Supplemental Figure 16. Feature analysis of task LANGUAGE. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task MOTOR)			
	Layer 1	Layer 2	Layer 3
(a)	54.62%	62.98%	89.00%

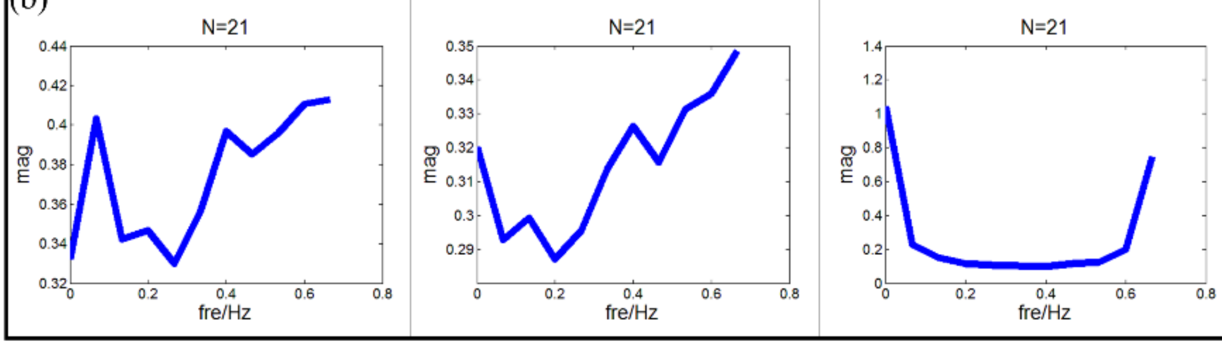
(b)



Supplemental Figure 17. Feature analysis of task MOTOR. (a) classification performance. (b) the distribution of frequency for features in average level.

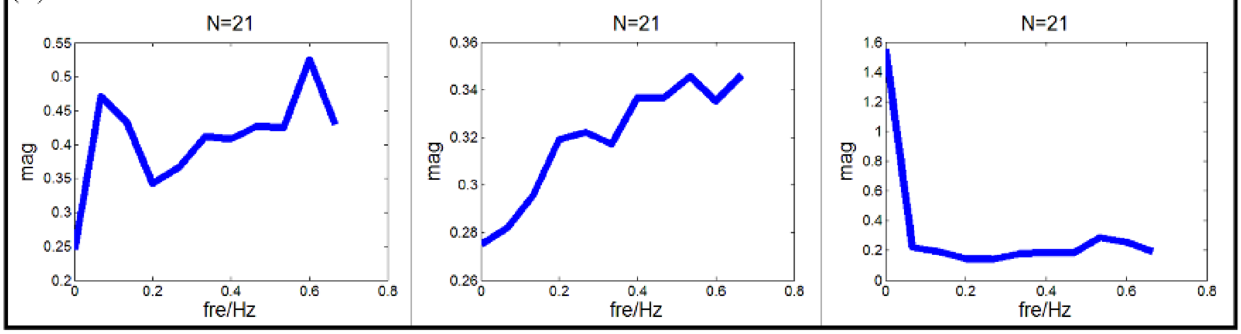
Performance of 1D-CNN architecture (Task RELATIONAL)			
	Layer 1	Layer 2	Layer 3
(a)	56.25%	63.59%	92.00%

(b)



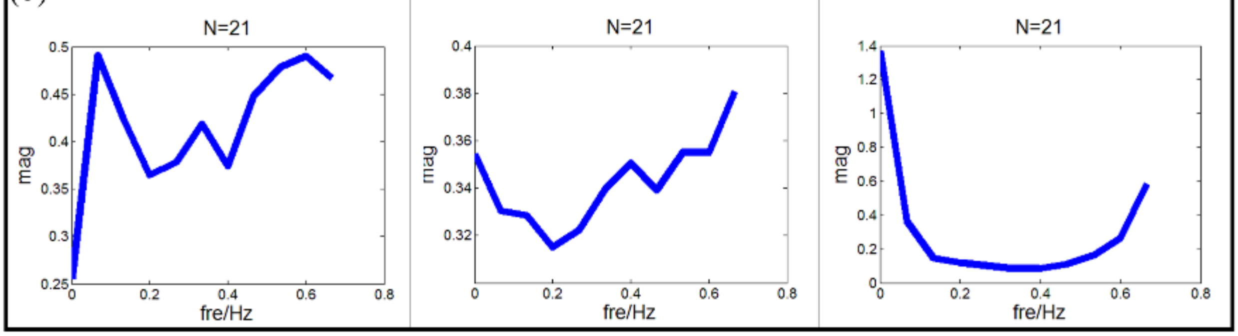
Supplemental Figure 18. Feature analysis of task RELATIONAL. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task SOCIAL)			
	Layer 1	Layer 2	Layer 3
(a)	55.09%	63.16%	90.95%
(b)			

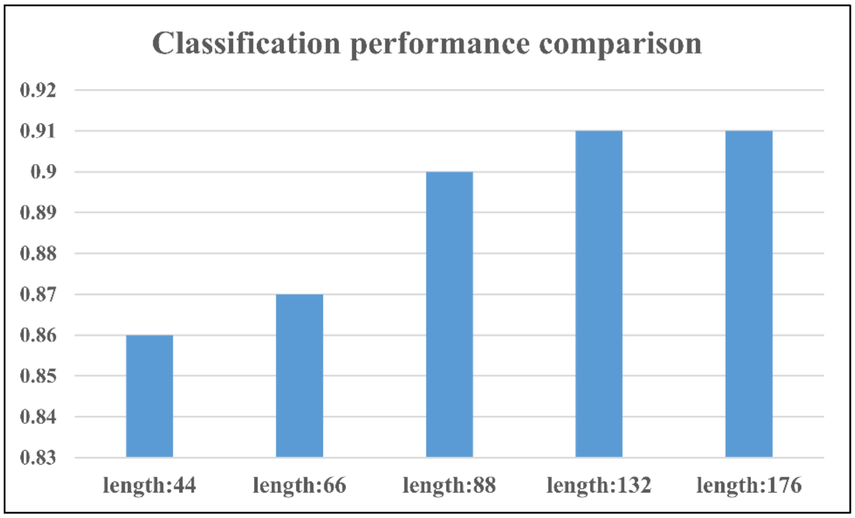


Supplemental Figure 19. Feature analysis of task SOCIAL. (a) classification performance. (b) the distribution of frequency for features in average level.

Performance of 1D-CNN architecture (Task WM)			
	Layer 1	Layer 2	Layer 3
(a)	52.50%	63.55%	88.26%
(b)			



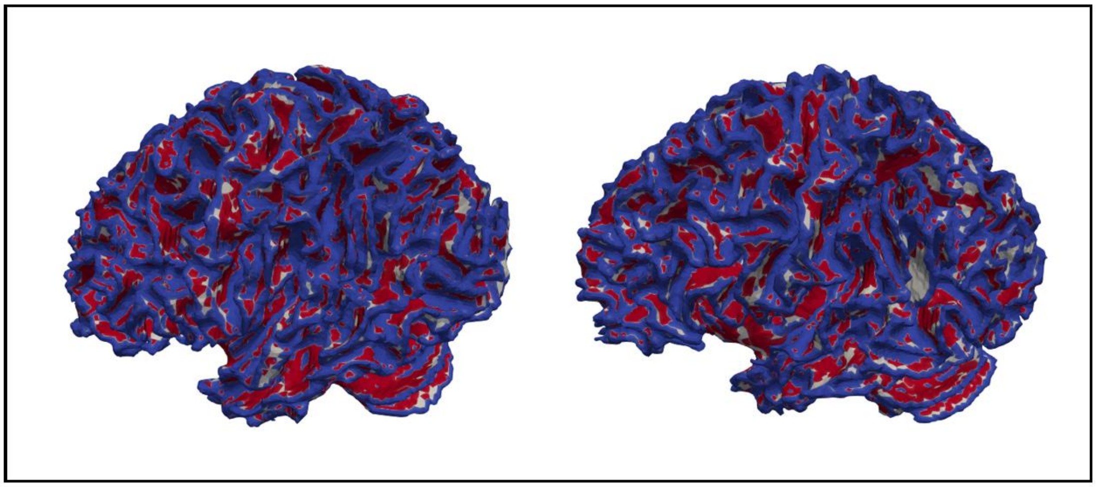
Supplemental Figure 20. Feature analysis of task WM. (a) classification performance. (b) the distribution of frequency for features in average level.



Supplemental Figure 21. Classification performance of middle vertices. Two examples (sbj1 and sbj2) are shown from left to right, color of red represents the vertices classified as gyri and color of gray represents the vertices classified as sulci.

Since in the training part, less time points will considerably reduce the training time, it is interesting to perform such test as the reviewer suggested. In the manuscript, we didn't try to change the length of fMRI signals in the training part. That is, in the CNN model, we used the original fMRI signals as the input for training.

However, as the review suggested, we tried to reduce the length of the input fMRI signals. We added 4 more experiments using different lengths of the fMRI signals as the training input and the results are summarized in the below paragraph. From the HCP dataset, we used task "EMOTION" as an example. The original length of the fMRI signal is 176. Then, we keep reducing the length of the time points and recorded the classification performances we learnt from the deep CNN models. As we can see from **Supplemental Figure 21**, when the length goes down to around 100 time points, the classification accuracy starts to drop. Similar results can be observed from other fMRI tasks. In general, the richer information we have in the input, the more accurate classification performance we can obtain. Reducing the length of the fMRI inputs will affect the classification accuracy of differentiating gyri/sulci, but it may not affect the classification performance too much when the length is over 100 time points. When the length is shorter than 100 time points, the classification performance will start to drop rapidly.



Supplemental Figure 22. Classification performance of middle vertices. Two examples (sbj1 and sbj2) are shown from left to right. Color of red represents the vertices classified as gyri and color of gray represents the vertices classified as sulci.

From the manuscript, we have already trained the CNN models using the vertices from both top gyri and deep sulci regions (excluding those middle vertices). Thus, we directly used those models to test the vertices with curvature values in the middle. In this way, the model keeps the best classification performance to distinguish the gyral and sulcal vertices, and it will help to classify vertices with curvature values in the middle into gyri and sulci clusters. The classification accuracy is about 60% for each subject, on average. That is, the vertices with curvature values in the middle are distinguishable, however, they are hard to be distinguished by using the features we already learned from top gyri and deep sulci. The reason we think is that fMRI signals of vertices in the middle are much more complex and they have correlations with both gyri and sulci.

Two classification examples are shown in the **Supplemental Figure 22**. As we can see, most middle vertices are classified as gyral vertices, maybe because gyri functions are more dominant in those regions. In other words, gyri functions are more global and thus it will affect much larger regions. In addition, we did notice that in some regions, middle vertices almost belong to one cluster (either gyri or sulci). The reason we think is that in those regions, function of gyri or sulci is dominant during the scan, but more details are to be examined in our future studies.

Supplemental Table 1 Testing classification accuracies of 6 groups from task EMOTION at the group level.

Group	1	2	3	4	5	6
Testing classification	**0.833**	**0.8453**	**0.818**	**0.813**	**0.842**	**0.783**