

Region-specific Brain Age Prediction Models for Children and Adolescents Derived by Machine Learning

Hyerin Jeong

iMediSync Inc

Ukeob Park

iMediSync Inc

Seung Wan Kang (✉ seungwkang@imedisync.com)

iMediSync Inc

Research Article

Keywords: Human brain, Prediction, Machine Learning, Brain age, Child and adolescent

Posted Date: August 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-839010/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

EEG biomarkers can reveal significant and actionable differences in brain development between normal children and those with developmental disorders. Frontal slow frequency EEG is one common differentiator between normal and abnormal brain function. The present study sought to establish models, based on machine learning, to predict brain age in children and adolescents. Four brain regions were studied: left anterior, right anterior, left posterior, and right posterior, based on the different functions characteristic of each region. Importantly, differences were also considered in the construction of the models. All models yielded promising r^2 values for the prediction of brain age, with values of 0.80 or higher. Our technique employed a tree-based feature selection algorithm, allowing selection of a minimum number of features while still preserving predictive power. These prediction models can be used to quantify deviations between estimated and biological brain age, and so serve as valuable tools in efforts to assess and intervene early in several profound developmental disorders.

1. Introduction

Puberty occurs in early adolescence, and stable emotional and intellectual development is critical during this period. Developmental disorders involving anxiety and other symptoms also commonly first arise during this period. The brains of children and adolescents are actively and dynamically growing neural circuits that regulate social behavior and other functions. However, in the later stages of this period, there is a gradual slowing of brain development, as brain growth is nearing completion [1].

EEG can serve as a biomarker of significant differences between normal children and adolescents and those with developmental disabilities ranging in severity from temporary and mild symptoms to ADHD and autism [2]. Previous EEG studies have investigated the onset of autism spectrum disorder (ASD) in newborns and epilepsy in infants [3, 4]. EEG has been established as a sensitive and non-invasive means of early detection and diagnosis that can lead to improved prevention and treatment of developmental disorders such as ADHD and ASD, which often first in early childhood. Several studies have shown significant correlations between specific EEG features and age [5, 6], and the brain maturation process of childhood and adolescence is clearly reflected in several EEG dynamics [7]. Importantly, recent studies have also revealed the importance of sex differences for the interpretation of EEG observations [8]. Additionally, sex differences specifically affect the probability of developing disorders such as anxiety, depression, psychosis, learning disability, and ADHD symptoms in childhood and adolescence [9]. Specific sex differences have been observed in relative power in the delta, theta, and alpha bands, and in the theta/alpha ratio (TAR) in normal children aged 8–12 years [10]. These sex differences in EEG patterns add complexity to those observed between normal and developmentally impaired disease groups. Significantly different EEG patterns have been well documented in the frontal region at delta, theta and slow alpha peak frequency between normal children and adolescents and those with developmental disabilities [11, 12]. For this reason, several studies have sought to predict age using EEG features [13, 14, 15].

Previous attempts to predict brain age to use MRI images and specific features extracted from the resting state EEG in both the eyes closed (EC) and eyes open (EO) conditions were limited by low subject numbers and poor brain region localization. Additionally, these studies were not specific to children and adolescents, which rendered it difficult to identify the specific EEG features affecting prediction for a specific age group. To overcome such limitations, we established four brain region-specific prediction models for both male and female groups comprising a dataset of 618 healthy children and adolescents. The four regions included the left anterior region, right anterior region, left posterior region, and right posterior region. The anterior region represents the frontal lobe, while the posterior region represents the rest of the brain covering temporal, parietal, and occipital lobe. All machine-learning models for each region were trained and validated using region-specific EEG features. This produced eight tree-based prediction models for each male and female group. We were able to identify EEG features that affect prediction results through specific calculation of their feature importance values, which were further used for feature reduction. As a result, all models returned promising r^2 values greater than 0.80.

2. Materials And Methods

2.1. Participants

Subjects were recruited through internet advertisements, and only healthy subjects were accepted into the study. The criteria for health included lack of neuropathic symptoms, lack of mental illness, and a normal mental and physical developmental history. After initial telephone-based screening for developmental and medical history, all subjects were further screened for depression and anxiety using self-administered questionnaires, while attention, memory, frontal function, and executive function were assessed using computerized neurocognitive tests. All procedures were approved by the Research Ethics Committee of the Seoul National University ((IRB number: 1711/003–004) and informed consent was obtained from each participant or guardian prior to enrollment. If the assessments of either depression or anxiety reached a clinical level, that subject was excluded. If any score on the cognitive domains was below $-2z$, or that on any three were below $-1.5z$, the subject was also excluded. Final enrollment included 618 healthy subjects aged 4 to 19 years. All methods were performed in accordance with the relevant guidelines and regulations.

In the metadata, the age groups were specified as two subgroups: a) children aged 4 to 6 years, and b) adolescents aged 6 to 19 years. Table 1 shows the number of subjects grouped by sex and age.

Table 1
The number of subjects in each sex and age group

Sex			
Age (years)	Male	Female	Total
a) 4 to 6	57	39	96
b) 6 to 19	245	277	522
Total	302	316	618

2.2. EEG recording and data processing

EEG was recorded using a Mitsar-EEG 202 digital electroencephalograph (Mitsar Ltd., St. Petersburg, Russia) with an electro cap (Electro-cap International, Inc., Eaton, Ohio, USA). Electrodes were placed on the surface of the scalp according to the international 10–20 system, at the following locations: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. The reference electrodes were placed on the mastoid process and the ground electrode was placed at Fpz. Electrode resistance was maintained below 5 k Ω . The EEG sampling rate was 250 Hz and the amplifier band ranged from 0.53 to 50 Hz. The participants were instructed to sit upright while the resting state EEG was measured under the eyes-open and eyes-closed conditions, for 3 minutes each. The raw EEG data were notch filtered using a low cut-off of 1 Hz and a high cut-off of 45 Hz. Re-referencing was performed using the common average reference (CAR) method. Artifacts were removed by bad epoch rejection and independent component analysis (ICA) using an automated, cloud-based QEEG analysis platform (iSyncBrain®, iMediSync Inc., Republic of Korea. <https://isyncbrain.com/>). Spectral power and power ratios were also calculated.

2.3. EEG Features for each region

Because the frontal brain area tends to undergo the greatest dynamic change during development [16, 17, 18], we divided the brain area into four regions for our analysis were comprised the left anterior, the right anterior, the left posterior, the right posterior areas.

Figure 1 shows the four areas designated for analysis. The features included in the analysis consisted of sensor level, source-level region of interest (ROI), and source imaginary coherence. Absolute power of the EEG was obtained from a fast Fourier transform (FFT) on each of the eight frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha1 (8–10 Hz), alpha2 (10–12 Hz), beta1 (12–15 Hz), beta 2 (15–20 Hz), beta3 (20–30 Hz), and gamma (30–45 Hz). Source level power was estimated by standardized low-resolution brain electromagnetic tomography (sLORETA), which includes 68 ROIs from the Desikan-Killiany atlas [19]. The imaginary part of coherence (iCoh), an indicator of brain connectivity [20] was calculated among each of 68 ROIs at eight frequency bands. All of the EEG features were analyzed using the automated cloud-based QEEG analysis platform (iSyncBrain®).

Sensor-level features included Fz, Cz, and Pz, which correspond to the center axis, in all four regions. T3 and T4 are not included in the frontal lobe but were excluded from the sensor-level features because they are too ambiguous to include in the posterior region. Features of the electrooculogram (EOG)-sensitive delta waves were only used to optimize noise reduction. The left anterior region included Fp1, F3, F7, Fz, C3, Cz; the right anterior included Fp2, Fp2, F4, F8, Fz, C4, Cz; the left posterior included C3, Cz, P3, Pz, T5, O1; and the right posterior included C4, Cz, P4, Pz, T6, O2.

2.4. Machine learning algorithm

We applied the random forest technique for feature selection, and then used feature importance, which can be determined in the order of predictive variables among all features. The random forest technique for ensemble models is based on a decision tree that can explain nonlinear relationships. Because these tree-based ensemble models have their own feature import attributes, they can be viewed by ordering variables according to their importance. In the present study, importance values were extracted based on the Scikit-learn algorithm. Only features with zero importance value were excluded, and the final model was established with enough features to secure predictive power while still maintaining the model's performance. Given the fact that there were up to 10,000 features, multiple feature selection processes were carried out while taking the variable importance of the machine learning model into account for maximized predictive power of the brain age prediction model.

The EEG dataset was split into a 4:1 ratio, for training and test datasets, respectively, and were validated through random five-fold cross-validation (cv). The male group comprised 241 training datasets and 61 test sets out of 302. The female group comprised 252 training datasets and 64 test sets out of 316. The predictive target variable was defined as the biological age.

In the process of developing the female and left posterior brain age prediction model, the cv score of the third fold was lower than that of the rest of the folds. To overcome the uneven scores of all five folds, cross-validation was performed 100 times for the same subjects for only the third fold, and the model was re-trained for all four regions.

The exclusion criteria for female subjects in this study were:

1. An absolute value of the difference between actual and predicted values of > 2 .
2. An absolute value of the difference > 2 for the same subjects only.

Thus, the brain age prediction model was finally developed for each of the four regions, with a total of 233 from the training set of 252 in the female group, excluding 19 subjects who met the above two exclusion criteria. Through this process, the prediction results of all five folds were evenly derived, thereby establishing a more stable prediction model.

3. Results

3.1. Performance of predictive models according to sex and region

We applied EEG features to predict brain age to regression models such as Lasso, Ridge, and ElasticNet, and regression models with tree-based machine learning techniques such as random forest, Xgboost, gradient boosting, and Lightgbm. The best model was then selected to derive the highest r^2 values based on the cross-validation set for each brain region. This study yielded promising r^2 values in all eight models for the prediction of brain age. The cross validation results are in Tables 2 and 3 below. Figure 2 and Fig. 3 present the eight scatterplots corresponding to all of the test dataset results of brain age prediction models. In the male group, the left anterior region model consisted of 26 features with an r^2 value of 0.90, the right anterior region model consisted of 57 features with an r^2 value of 0.91, the left posterior region model consisted of 38 features with an r^2 value of 0.84, and the right posterior region model consisted of 41 features with an r^2 value of 0.87. In the female group, the left anterior region model consisted of 36 features with an r^2 value of 0.88, the right anterior region model consisted of 37 features with an r^2 value of 0.88, the left posterior region model consisted of 46 features with an r^2 value of 0.87, and the right posterior region model consisted of 45 features with an r^2 value of 0.87.

Table 2

Cross-validation results of four different regression models for males and corresponding brain region

Region	ML algorithm	CV scores (r^2)
left anterior	random forest	0.89
right anterior	random forest	0.88
left posterior	random forest	0.87
right posterior	random forest	0.90

Table 3

Cross-validation results of four different regression models for females and brain corresponding region

Region	ML algorithm	CV scores (r^2)
left anterior	random forest	0.90
right anterior	Xgboost	0.89
left posterior	random forest	0.87
right posterior	random forest	0.98

In addition, to show sex differences, female datasets were constructed by switching male models with female features for each region, and vice versa: male datasets were constructed by changing female models to male features for each region. As a result, test scores were significantly reduced when testing datasets were replaced with that of the opposite sex in each sex-specific predictive model. Since there is a sex difference in the EEG of growing children and adolescents, it can be significant that predictive models have been established by dividing sex.

3.2. Important features selected by ML techniques

We confirmed in all brain age prediction models that the delta and theta at Fz, Cz, Pz has a higher feature importance value. This shows that brain age is relatively influential in predicting these features compared other frequencies and channels. Based on the finding above, we investigated differences in EEG between a patient with developmental disabilities and normal child. EEG data from individuals excluded from the above analysis due to suspected developmental problems were used. Figure 4 shows the difference in EEG band power between a normal child and one with suspected developmental problems using the iSyncBrain® 1:1 comparison analysis. Both subjects were 6.1 years old. In Fig. 4, G1 and G2 are the child with suspected developmental problems and the normal child based on the normative database, respectively. Since the differences between them were mostly negative, this suggests that the relative power of the delta and theta bands are likely higher for subjects showing symptoms of possible developmental problems. This group comparison also verified that there are significant differences especially in delta and theta in the Fz, Cz, and Pz channels.

4. Discussion

Exponential physical and mental development takes place during childhood and adolescence, and it is therefore essential to carefully assess emergent symptoms of anxiety and developmental disorders during this critical period. Previous studies have sought a means of predicting developmental disorders by analyzing the brain's connectivity using MRI and EEG analyses. However, EEG has several advantages over MRI, including lower costs, portability, and better temporal resolution, while offering similar diagnostic power. EEG data gains its greatest utility for disease prediction when referenced to a well-constructed normative database. For example, the EEG of children with developmental disorders demonstrate significant differences in the delta and theta at the frontal lobe compared with those of normal children.

The present findings are consistent with prior findings that the brain wave patterns of children and adolescents are particularly dynamic. All age prediction models specific to brain regions and sex developed in the present study derived high r^2 values. Such results demonstrate the utility and cost-effectiveness of assessing the awake and resting state EEG and the potential of such models to predict specific disorders based on region-specific comparisons. The reliability of brain age prediction models increases with as the number of children and adolescents tested increase. The present analysis is unique in that the predictive values were derived for each of the four regions of the brain divided by that for the

frontal region. Figure 1 shows the sensor level features belonging to each of the four brain regions. Fz, Cz, and Pz data were included in the analysis of each region. Figures 2 and 3 present scatterplots of all prediction models for each test set. They reveal significant linear correlations between the actual biological age and the predicted value of each brain age prediction model.

The predictive models have high predictive power utility and are parsimonious in that the minimum number of features can be used while preserving their predictive power by applying the tree-based feature importance technique. The Fz, Cz, and Pz feature importance of delta and theta are derived at the top rank in all brain age prediction models by having larger values than other features, and these features are prominent in Fig. 4. An important factor in the present strategy is the use of region-specific investigation using machine learning techniques trained exclusively on normal subjects aged 4 to 19 years. This enabled prediction of developmental problems based on region-specific machine learning algorithms validated by actual biological age of a subject and a predicted value for each of the four regions.

Future studies will extend these ROI-specific, anatomically relevant brain age prediction algorithms by including eight different brain lobal areas among the frontal, temporal, parietal, and occipital lobes in both hemispheres. Furthermore, since the distribution of predicted values varies markedly with age, we will attempt to identify acceptable ranges of predicted ages within specific age groups. Through iterative refinement powered by additional data, the brain age prediction models established in the present study will continually improve and show greater predictive power and validity.

5. Conclusion

EEG patterns change markedly and dynamically during early development and are sex-specific. The present study established a brain age prediction model using EEG feature-based machine learning techniques applied to EEG features from normal children and adolescents exclusively. Many previous neurological studies have employed EEG analysis of children with developmental disabilities. However, the present study is the first to limit the age range and focus solely on normal children and adolescents with additional attention to the variables of sex and region to derive region-specific age prediction models. There is general consensus on the value of early detection and intervention in developmental disorders. The present findings provide predictive values for each of the four lobes, which may contribute to efforts to detect, interpret, and treat these disorders informed by the specific functions of each lobe. Our brain region-, age-, and sex-specific approach yielded eight brain age prediction models with high r^2 values, which suggests these models offer great promise for reducing the suffering the accompanies these challenging developmental disorders.

Declarations

Acknowledgements

This research was supported by the Research and Business Development Program through the Korea Institute for Advancement of Technology (KIAT) funded by the Ministry of Trade, Industry and Energy (MOTIE) (grant number: P0014055), and by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) and Korea Dementia Research Center (KDRC), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (grant number: HU20C0511000020)

Author contributions

H.R.J., U.E.P., and S.W.K. organized the project. H.R.J., U.E.P., S.W.K. performed data analysis and visualization. H.R.J., U.E.P., and S.W.K. drafted the manuscript text. S.W.K. led the project and oversaw manuscript preparation.

References

1. Schulz KM, Sisk CL. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neurosci Biobehav Rev.* 2016;70:148-158. doi:10.1016/j.neubiorev.2016.07.036
2. Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and Anxiety Disorders in Children and Adolescents: Developmental Issues and Implications for DSM-V. *Psychiatric Clinics of North America*, 32(3), 483–524. <https://doi.org/10.1016/j.psc.2009.06.002>
3. Bosl, W. J., Tager-Flusberg, H., & Nelson, C. A. (2018). EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-24318-x>
4. van Diessen, E., Otte, W. M., Braun, K. P. J., Stam, C. J., & Jansen, F. E. (2013). Improved Diagnosis in Children with Partial Epilepsy Using a Multivariable Prediction Model Based on EEG Network Characteristics. *PLoS ONE*, 8(4), e59764. <https://doi.org/10.1371/journal.pone.0059764>
5. Matthis, P., Scheffner, D., Benninger, Chr., Lipinski, Chr., & Stolzis, L. (1980). Changes in the background activity of the electroencephalogram according to age. *Electroencephalography and Clinical Neurophysiology*, 49(5–6), 626–635. [https://doi.org/10.1016/0013-4694\(80\)90403-4](https://doi.org/10.1016/0013-4694(80)90403-4)
6. Marshall, P. J., Bar-Haim, Y., & Fox, N. A. (2002). Development of the EEG from 5 months to 4 years of age. *Clinical Neurophysiology*, 113(8), 1199–1208. [https://doi.org/10.1016/s1388-2457\(02\)00163-3](https://doi.org/10.1016/s1388-2457(02)00163-3)
7. Cragg, L., Kovacevic, N., McIntosh, A. R., Poulsen, C., Martinu, K., Leonard, G., & Paus, T. (2011). Maturation of EEG power spectra in early adolescence: a longitudinal study. *Developmental Science*, 14(5), 935–943. <https://doi.org/10.1111/j.1467-7687.2010.01031.x>
8. Karlsgodt, K. H., John, M., Ikuta, T., Rigoard, P., Peters, B. D., Derosse, P., Malhotra, A. K., & Szeszko, P. R. (2015). The accumbens-frontal tract: Diffusion tensor imaging characterization and developmental change from childhood to adulthood. *Human Brain Mapping*, 36(12), 4954–4963. <https://doi.org/10.1002/hbm.22989>

9. Kaczurkin, A. N., Raznahan, A., & Satterthwaite, T. D. (2018). Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology*, 44(1), 71–85. <https://doi.org/10.1038/s41386-018-0111-z>
10. Clarke, A. R., Barry, R. J., McCarthy, R., & Selikowitz, M. (2001). Age and sex effects in the EEG: development of the normal child. *Clinical Neurophysiology*, 112(5), 806–814. [https://doi.org/10.1016/s1388-2457\(01\)00488-6](https://doi.org/10.1016/s1388-2457(01)00488-6)
11. Arns, M., Gunkelman, J., Breteler, M., & Spronk, D. (2008). EEG phenotypes predict treatment outcome to stimulants in children with ADHD. *Journal of integrative neuroscience*, 7(03), 421-438. <https://doi.org/10.1142/s0219635208001897>
12. Kanemura, H., Sano, F., Tando, T., Sugita, K., & Aihara, M. (2013). Can EEG characteristics predict development of epilepsy in autistic children? *European Journal of Paediatric Neurology*, 17(3), 232–237. <https://doi.org/10.1016/j.ejpn.2012.10.002>
13. Dimitriadis, S. I., & Salis, C. I. (2017). Mining Time-Resolved Functional Brain Graphs to an EEG-Based Chronnectomic Brain Aged Index (CBAI). *Frontiers in Human Neuroscience*, 11. <https://doi.org/10.3389/fnhum.2017.00423>
14. Al vandenbosch, O., Ki Wong, C., Kuplicki, R. T., Yeh, H., Mayeli, A., Refai, H., Paulus, M., & Bodurka, J. (2018). Predicting Age From Brain EEG Signals—A Machine Learning Approach. *Frontiers in Aging Neuroscience*, 10. <https://doi.org/10.3389/fnagi.2018.00184>
15. Vandenbosch, M. M. L. J. Z., van 't Ent, D., Boomsma, D. I., Anokhin, A. P., & Smit, D. J. A. (2019). EEG-based age-prediction models as stable and heritable indicators of brain maturational level in children and adolescents. *Human Brain Mapping*, 40(6), 1919–1926. <https://doi.org/10.1002/hbm.24501>
16. Sowell, E. R., Delis, D., Stiles, J., & Jernigan, T. L. (2001). Improved memory functioning and frontal lobe maturation between childhood and adolescence: a structural MRI study. *Journal of the International Neuropsychological Society*, 7(3), 312-322. <https://doi.org/10.1017/s135561770173305x>
17. Whitford, T. J., Rennie, C. J., Grieve, S. M., Clark, C. R., Gordon, E., & Williams, L. M. (2006). Brain maturation in adolescence: Concurrent changes in neuroanatomy and neurophysiology. *Human Brain Mapping*, 28(3), 228–237. <https://doi.org/10.1002/hbm.20273>
18. Segalowitz, S. J., & Davies, P. L. (2004). Charting the maturation of the frontal lobe: An electrophysiological strategy. *Brain and Cognition*, 55(1), 116–133. [https://doi.org/10.1016/s0278-2626\(03\)00283-5](https://doi.org/10.1016/s0278-2626(03)00283-5)
19. Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
20. Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10), 2292–2307. <https://doi.org/10.1016/j.clinph.2004.04.029>

Figures

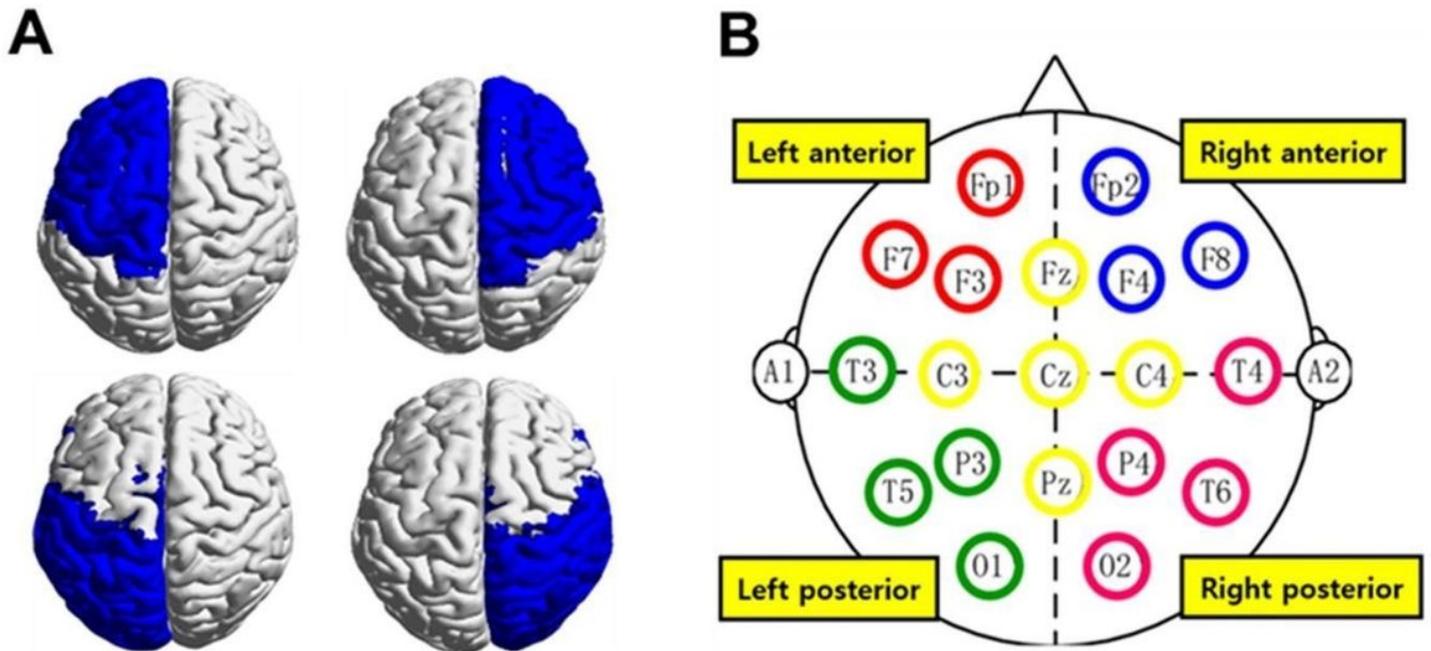


Figure 1

A) Four brain regions: left anterior, right anterior, left posterior, right posterior. B) Montage of international 10-20 system for four brain regions.

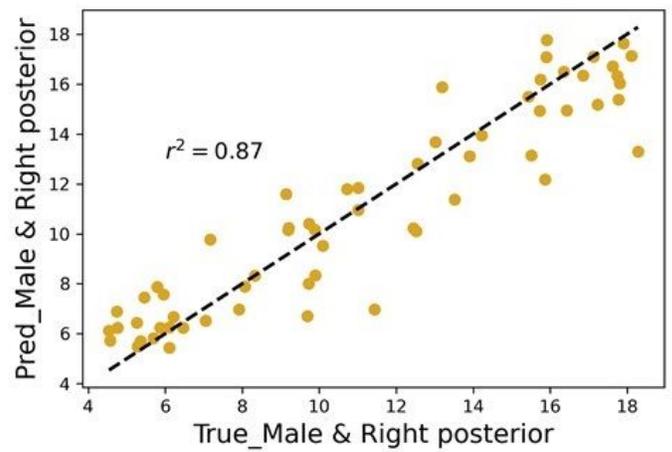
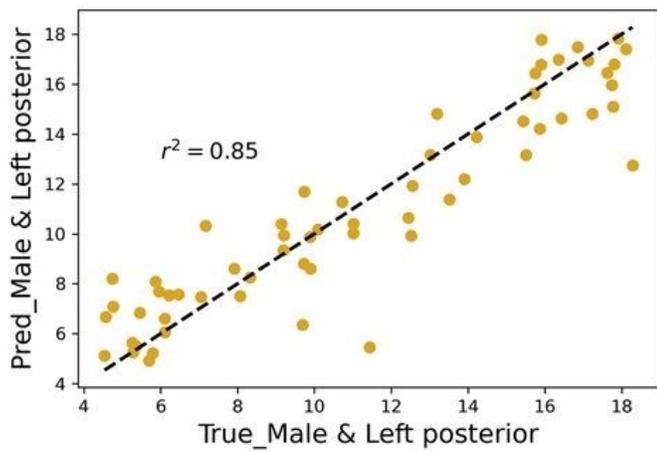
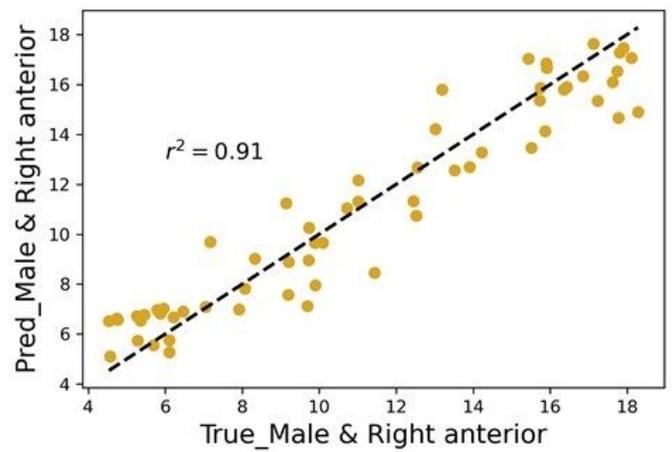
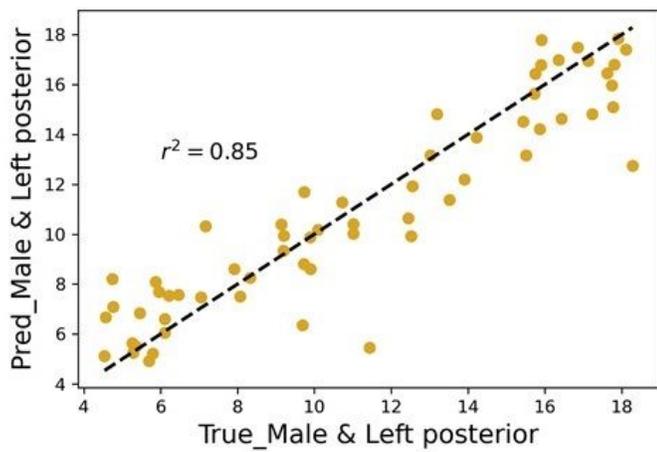


Figure 2

Scatterplot predicts test dataset with brain age prediction models for males by region.

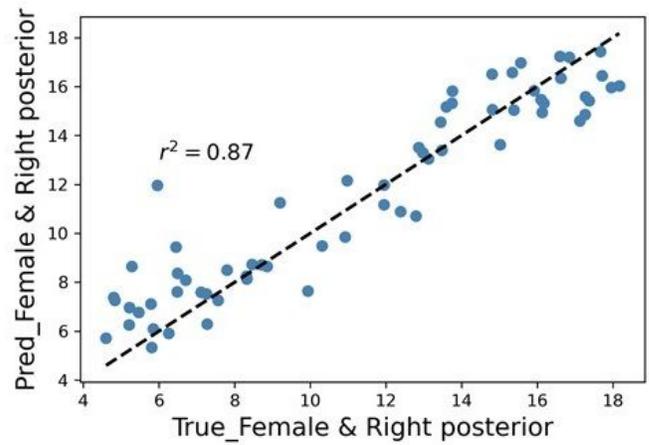
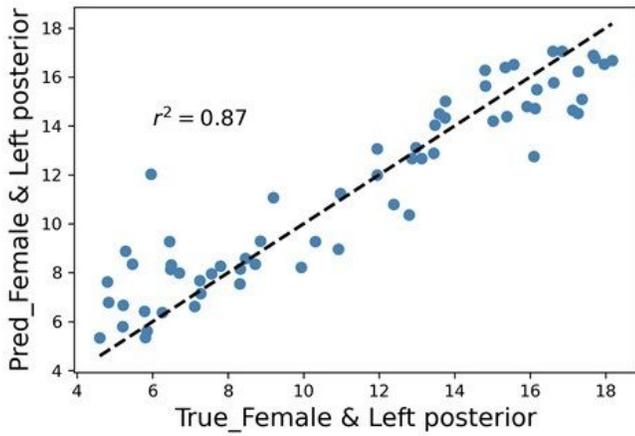
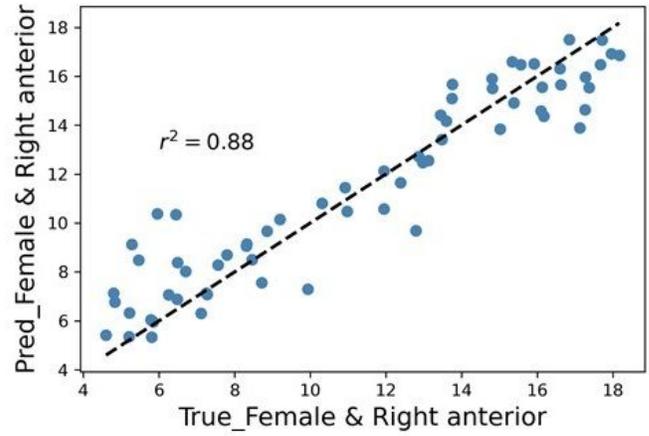
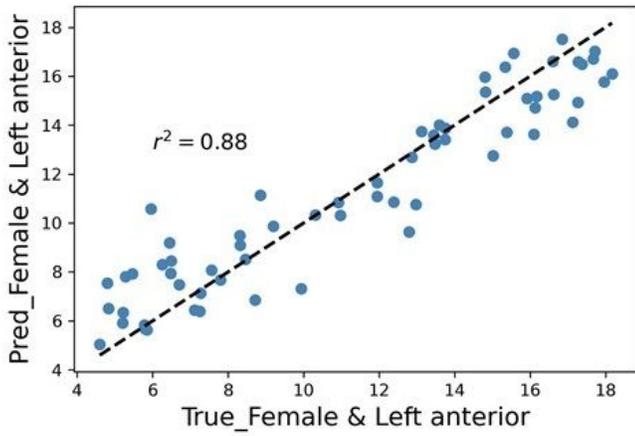
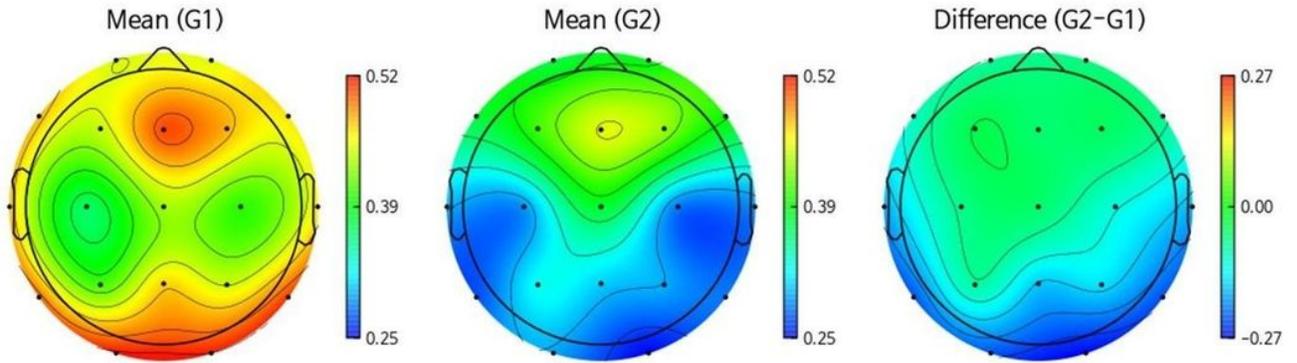


Figure 3

Scatterplot predicts test dataset with brain age prediction models for females by region.

[Topomap (Rel. power) - Delta band]

Unit: %



[Topomap (Rel. power) - Theta band]

Unit: %

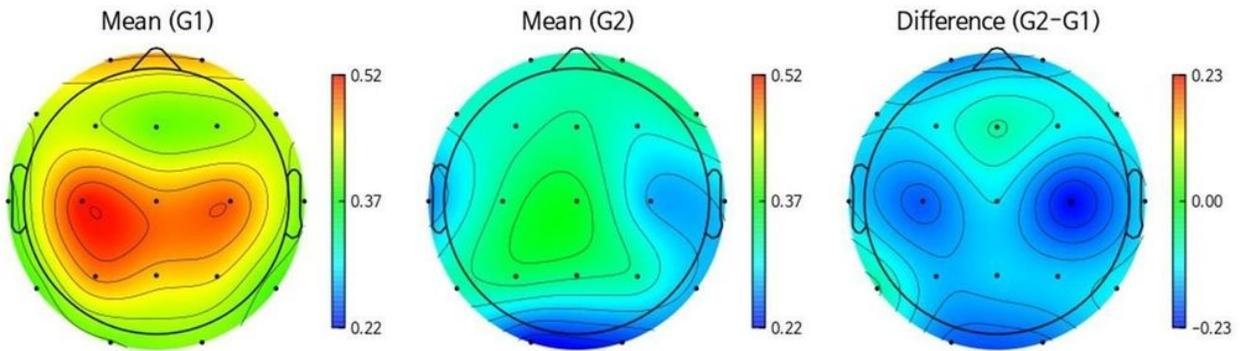


Figure 4

Delta/theta band relative power comparison between normative database and a subject showing symptoms of developmental disorder.