Data Science Lab Report

MRI and EEG data for brain age and psychiatric disorders prediction

Camilla Casamento Tumeo
ETH Zürich
camillca@ethz.ch

Emanuele Palumbo
ETH Zürich
palumboe@ethz.ch

Alessandro Stolfo
ETH Zürich
stolfoa@ethz.ch

Abstract

Structural brain information can nowadays easily be collected with relatively non-invasive procedures. This project investigates to what extent structural brain data, in the form of features extracted from magnetic resonance imaging (MRI) and electroencephalography (EEG), can be informative on the mental health of children and adolescents. We worked on data from a pediatric sample of 2096 patients, who may be either healthy or diagnosed with one (or more) psychiatric morbidities. In a first phase of the project, we attempted to use MRI and EEG data to explain the variation in both behavioural (SWAN score [8], WISC score [11]) and biological (Age) measures. Age prediction gave the most promising results. The age predicted from MRI [5] and EEG [1] data can be interpreted as the "brain age" of the patient. While it is natural to assume that such brain age reflects well the actual (anagraphic) age in a healthy subject, we explored whether this is the case for patients affected by psychiatric morbidities. Thus, we focused on the prediction of the brain age of an individual, using a conditional density estimation (CDE) model [9]. We then analyzed the patients' brain-age discrepancy [10], which is the difference between the predicted brain age and the anagraphic age of an individual, and has already been related to the health of the subject ([4]), where a large discrepancy can be linked to reduced fitness or health-outcomes. Finally, we showed how a significant difference in the size of the brain-age discrepancy can be identified in the patients diagnosed with autism, when compared to healthy individuals.

1 Introduction and Objectives

In the recent years technical developments have boosted the usage of structural brain information for research in cognitive neurosciences and clinical psychiatry. This is mostly due to the high quality and relatively non-invasive nature of the acquisition methods that are nowadays available. Examples of these methods are Electroencephalography (EEG), structural Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI). EEG [2] records brain activity by measuring electrical potential with electrodes that are placed on the scalp: EEG data have already [6] been proved to be useful in diagnosing psychiatric morbidities. Structural MRI gives information on grey and white matter in the brain, in terms of different indicators (e.g. shape, integrity). DTI detects fiber structure and maturity of white matter in the brain. [7] gives an overview on these neuroimaging techniques and on their relevance in cognitive neurosciences, especially when dealing with young patients.

The aim of this project is to understand in what measure structural brain information, in the form of MRI and EEG data, can be informative on the mental health of children and adolescents. Initially, we directly attempted to predict the presence of specific disorders from EEG data with standard methods, but EEG data seemed not to contain enough signal to obtain a reasonable accuracy in this

task. Results relative to this are reported in section 7. We then drifted towards the more reasonable goal of predicting some behavioral/anagraphic scores from MRI/EEG. While prediction of behavioral scores (see section 6) was not promising, we obtained interesting results in age prediction. From these results, we developed our research further, interpreting the predicted age of a patient as his brain age ([1],[5]), i.e. the age that can be inferred from the patient's brain structure. Assuming that brain age reflects well the anagraphic age for a healthy patient, we trained a conditional density estimation model for age on healthy patients only, using MRI features. Then, we conducted an analysis to investigate if a large discrepancy between predicted brain age and actual age of a patient can be informative of the patient's psychiatric condition. Interestingly, patients affected by autism showed a significantly higher brain age discrepancy when compared to healthy patients, as will be discussed in section 5.

2 Dataset Description

Our dataset is composed of three parts: a behavioral dataset, an MRI dataset and an EEG dataset. The total number of patients in our behavioral dataset is 2096. EEG data and MRI data are both available only for a subset of the total patients. The amount of patients for which we have the three kinds of data available is around 800.

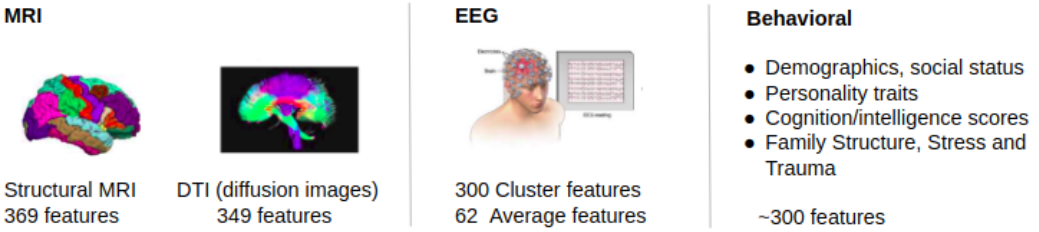

Figure 1: Available data

2.1 Behavioral dataset

The behavioral dataset records, for each patient, the following information:

- **Psychiatric disorders** a patient is affected by.
  For each patient up to 10 psychiatric disorders are reported, ordered by their severeness. For every disorder the dataset reports the general category of the disorder (e.g. "Neurodevelopmental Disorder"), a subcategory (e.g. "Attention-Deficit/Hyperactivity Disorder") and the specific name of the disorder (e.g. "ADHD-Inattentive Type"). These diagnoses are the result of a medical evaluation performed by doctors.
- **Behavioral scores** (about 270 features).
  Many of these scores are results of questionnaires taken by the patient and reveal aspects about his family structure, cognitive capabilities (e.g. WISC score), personality traits, general behavioral attitudes (e.g. SWAN score). Information as social status, age and gender is also present.

The behavioral dataset contains missing values for the most part: more than 50% of its features have more than 90% of missing values indeed. The presence of a missing value is in this context a meaningful information itself, since patients were assigned questionnaires based on how they answered to previous questionnaires.

2.2 MRI dataset

The MRI dataset contains two sets of features:

- **Structural MRI**
  The structural MRI dataset contains 369 features, which are available for 1146 patients in total. We did not work with raw MRI images, but rather with already extracted features,

representing thicknesses and volumes of different parts of the brain in the cortical and sub-cortical regions. These data are extracted for both the left and the right emisphere.

- **DTI (Diffusion Tensor Imaging)**
  This part consists of 349 extracted features, which are related to the diffusion of water molecules inside the brain. Molecular diffusion in tissues is not free, but reflects interactions with many obstacles, such as macromolecules, fibers, and membranes. Water molecule diffusion patterns can for example help identifying white matter regions inside the brain. These features thus contain information about the tissue composition of the brain. We have a total of 838 patients having both structural MRI and DTI data.

2.3 EEG dataset

The dataset is based on data from 1485 subjects from two experimental conditions: eyesclosed (eyes-closed resting-state EEG recordings, 40-sec duration) and eyesopen (eyes-open resting-state EEG, 20-sec duration).
The original EEG data was preprocessed with the Automagic toolbox [1].
We worked with the **Spectrogram-based** features of the preprocessed dataset. The dataset contains three levels of Spectro features:

- Channel: features for each of 105 EEG channels (5054 features)
- Cluster: features for each of 6 EEG channel clusters (302 features)
- Average: averaged features across all EEG channels (62 features)

For our experiments, we used only average and cluster features.

Note: Both MRI and EEG exams were taken on patients in resting state, meaning that the patient was *not* answering to any questionnaire of the behavioral dataset while his MRI scans/EEG signals were registered.

3 Age Prediction

In this section, we show that it is possible to obtain interesting results in age prediction from MRI/EEG data. We predict age using MRI features only, EEG only and both. The resulting 5-fold cross validation and test MSEs are shown in Table 4. As a baseline to compare with, we use the sample mean of the patients' age, which gives an MSE of about 13. Linear models as Lasso do not perform well in this task, suggesting a more complex relation between the features and the dependent variable. While SVR with *rbf* kernel and tree-based feature selection improves considerably over the simpler linear models, our best results are obtained with XGBoost and CDE (conditional density estimation neural network, which will be described in higher detail in the next section). As can be seen from Table 4, the MSEs obtained with these models are around 5 when training on MRI only and slightly worse when training on EEG only; the error drops to around 4 when using both sets of features combined, meaning that information coming from MRI and EEG is not redundant.

Table 1: MSEs for Age prediction

	MRI		EEG		MRI + EEG	
	5-CV MSE	Test MSE	5-CV MSE	Test MSE	5-CV MSE	Test MSE
Sample Mean	–	13.82	–	12.91	–	12.97
Lasso	10.05	9.88	8.25	8.30	7.70	7.43
XGBoost	4.83	4.92	5.10	4.83	**3.78**	3.99
SVR	4.97	5.17	5.25	5.05	4.29	4.43
NN-based CDE	4.92	5.46	5.27	6.34	4.01	**3.89**

When working with MRI we tried training with structural MRI only and with structural MRI and DTI together. Adding the DTI features did non bring any improvement, possibly also due to the fact that

[1] `https://github.com/methlabUZH/automagic`

DTI are available for a restricted subset of patients, thus reducing the number of training samples. The results reported in the table have been obtained using structural MRI only, without DTI. When working with EEG we repeated our experiments with average features and cluster features and we observed slightly better results when using cluster features, which contain more information: also in this case, the EEG results reported in the table are the best we obtained, meaning that they are relative to models trained on EEG *cluster* features. All the models we employed have been fine-tuned with grid-search. The fine-tuned hyperparameters of our best performing model (XGBoost), trained on MRI and EEG together are *regression_booster_max_depth = 3* and *regression_booster_alpha = 0*. The EEG missing values were imputed with the median.

4 Conditional Density Estimation

One of the models we employed in order to predict age from MRI/EEG features is a conditional density estimation model with noise regularization, following the same approach as [9]. In this case, the model does not only output an estimate of the age, but the whole age conditional probability distribution is modeled; the distribution parameters are outputted by a neural network. As a density model, we employ a Gaussian Mixture Model (GMM) with diagonal covariance matrices. The conditional density estimate of the age becomes

$$\widehat{p}(y|x) = \sum_{k=1}^{K} w_k(x,\theta)\mathcal{N}(y|\mu_k(x,\theta), \sigma_k^2(x,\theta))$$

being $w_k(x,\theta), \mu_k(x,\theta), \sigma_k^2(x,\theta)$ the weight, mean and variance of the k-th Gaussian component of the GMM. These parameters are outputted by the network, while K, the number of Gaussian components, is chosen to be 10. The employed architecture is a simple neural network with two hidden layers of 16 nodes each. A soft-max activation function is used on the output neurons corresponding to the Gaussian weights, in order to enforce their sum to be one, while a softplus activation function is used on the variance output neurons, to make them positive. Since our dataset is relatively small, we employ noise regularization during training, as proposed in [9]. The idea of noise regularization is to perturb the training data in every mini-batch with noise sampled from a Gaussian with zero mean and variance $h \in \mathbb{R}$. It was shown in [9] that adding such a perturbation to the data in the context of conditional density estimation for small datasets has a regularizing effect on the shape of the estimated distribution, leading in general to better results. To choose h appropriately, we used the rule-of-thumb value that proved to performed well in different contexts in [9], i.e. $h = n^{-\frac{1}{d+1}}$, where n is the number of observations in the dataset and d is the sum of the dimensions of the feature space and target space (the target space dimension is 1 in our case, since our goal is age prediction).

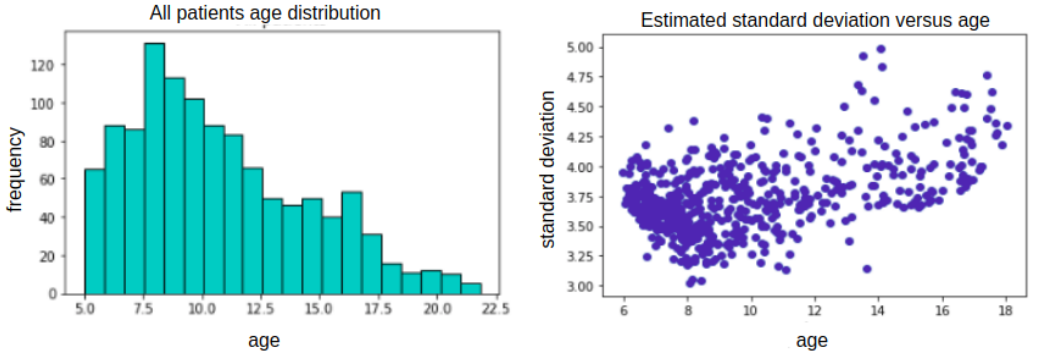

Figure 2: Age distribution in the whole population (left) and CDE estimated standard deviation versus age (right).

As figure 2 shows, the CDE model is capable of correctly estimating the standard deviation of the distribution: such uncertainty indeed increases for older patients, which reflects the left-skewed age distribution of the patients (we have less observations for older patients). The advantage of employing such a density estimation model, which gives us much more information than punctual predictions, will be more evident in the next section, when we will define a new feature, namely the "brain age discrepancy" for each patient.

5 Brain Age Discrepancy

We showed that it is possible to predict the age of an individual with reasonable accuracy from MRI data. Such prediction can be interpreted as the "brain age" of a patient [5; 1], i.e. the age that can be inferred from his brain structure. It is natural to expect the brain age of a healthy person to be the same as his actual age. Thus, it becomes interesting to investigate if a discrepancy between the actual age of a patient and his predicted brain age (from MRI) may be correlated with the patient main psychiatric disorder. This is what we take care of in this section.

5.1 Definition of brain-age Δ

We introduce the new feature "brain-age Δ" (or "brain-age discrepancy") for the patients. In order to define such a feature, we make use of the conditional density estimation model we introduced in the previous section. As we said, such a model is able to predict not only the mean, but also the standard deviation (i.e. uncertainty) of the patient age distribution given his MRI data. We exploit this information in order to define the brain-age discrepancy for a patient. Given the MRI data of a patient, we use the CDE model to predict age_{brain}, i.e. the predicted mean of the age distribution of that patient , and σ, the corresponding standard deviation. If age is the real age of the patient, we define the patient brain-age discrepancy as

$$\Delta = \frac{age - age_{brain}}{1 + \sigma}$$

In this way, the brain-age discrepancy also contains a measure of the uncertainty in our prediction: to obtain a high brain-age discrepancy for a patient we not only need his actual age to be far from his predicted brain age, but also that the predicted brain age variability of that patient is not too high. The choice of adding 1 to σ in the denominator comes from the need of controlling the size of the ratio when σ becomes small. In other words, we do not want Δ to become too large because the uncertainty about the prediction of the brain age is small.

5.2 Experiment setting

Our goal is to investigate if there is a correlation between a high/low brain-age discrepancy and the psychiatric disorder of a patient. We conduct this analysis for the four most frequent disorders in our dataset, i.e.

- Autism
- ADHD-Inattentive
- ADHD-Combined
- Specific Learning Disorder with Impairment in Reading

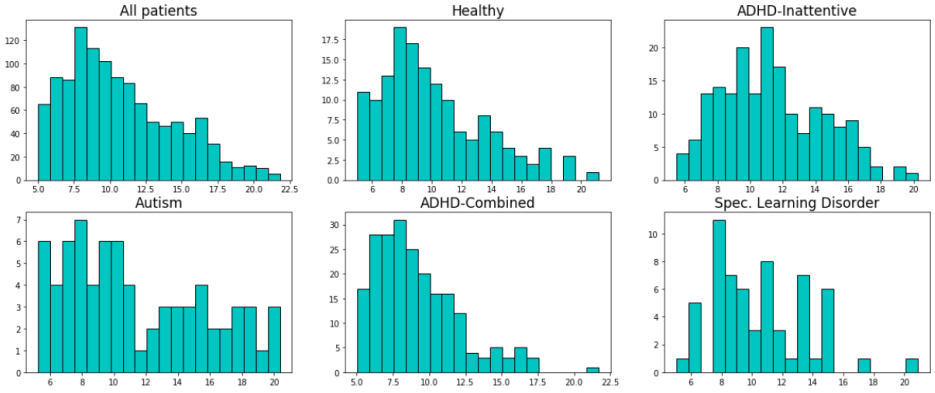

Figure 3: Age distribution in the whole population (top left) and in the subpopulations of healthy (top center) and disturbed patients.

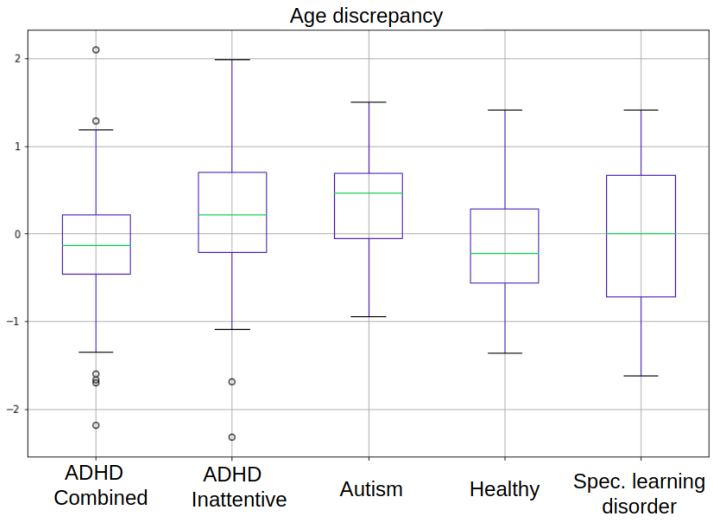

Figure 4: Boxplot of Δ for each morbidity class.

Figure 3 depicts the age distributions for all the patients, healthy patients, and patients affected by the four disorders of interest. The dataset we use to predict age distribution is MRI only, since it proved to give good results as illustrated in the previous section. We do not use MRI and EEG fetures together since the subset of patients having both sets of features is rather restricted and we do not want to lose data samples in this hypothesis testing part of the project. Starting with the assumption that actual age and brain age should be the same for healthy patients, we train our CDE model on half of the healthy patients. The CDE noise regularization is crucial here, since the model still performs well even when training on few observations. In this experiment, the training set is composed of a half of the healthy patients (74 observations). The "test set", i.e. the set we perform our further analysis on is composed of the other half of the healthy patients, plus patients affected by any of the four major disorders we are taking into consideration. We use the model trained on healthy patients only to predict brain age discrepancy for all patients in the test set. We then perform four t-tests on the test set, to see if the mean of the brain age discrepancy for patients with psychiatric disorders is significantly different from the mean of the brain age discrepancy of healthy patients. The p-values are corrected for multiple testing using the Benjamini-Hochberg criterion. The procedure is repeated over 5 different splits of the healthy patients. The first columns of Table 2 reports the corrected p-values, averaged over the different splits and the range of minimum-maximum p-values found in the 5 different splits, while Fig. 4 shows a boxplot of the brain age discrepancy grouped by psychiatric disorder for a random split.

Table 2

	t-test p-val		p-val controlling for age	
ADHD-Inattentive	0.03%	(0.00%-0.13%)	6.88%	(0.03%-24.90%)
ADHD-Combined	60.45%	(22.18%-76.11%)	12.81%	(0.72%-37.41%)
Autism	0.10%	(0.01%-0.30%)	**4.00%**	(0.05%-13.96%)
Spec. Learning Disorder	4.23%	(1.32%-9.19%)	9.62%	(0.72%-29.56%)

Although this first test gives us important information about the difference in age discrepancy among ill and healthy patients, it is important to repeat the experiment while controlling for age. Indeed, there is a positive correlation between age and age discrepancy, partially due to the fact that we have few observations for older patients. It is thus important to control for age, to be sure that age is not a confounding variable that justifies the significant differences in age discrepancy of the different classes of ill patients. In this experiment, we simply compute the p-values of a linear model in which the dependent variable is brain age discrepancy and the predictors are age and the disorder. As before, the p-values are corrected for multiple testing and averaged over 5 possible different splits of the

healthy patients. The results are reported in the second column of Table 2. The disorder category "healthy" was used as a baseline and the p-values are relative to the four major disorders when comparing with the baseline and controlling for age.

It is interesting to notice that some p-values that were significant before are not significant anymore when controlling for age. Nevertheless, we obtain a significant p-value for autism at significance level $\alpha = 0.5\%$. Though it is hard to infer causal relationship in this context, this result interestingly shows that there may be a correlation between having a certain disorder (in this case, autism) and presenting a brain age discrepancy higher than normal. This would mean that patients with autism have a brain that looks "younger" (i.e. less developed) than it should be, and such information could be helpful when diagnosing specific psychiatric disorders.

6 SWAN Score and WISC Score Prediction

Together with the work on the age prediction, we also tried to model the variation of the scores of two behavioural tests:

- **SWAN score**
  SWAN, which stands "Strengths and Weaknesses of Attention-Deficit/Hyperactivity-symptoms and Normal-behaviors" [8], is a rating scale meant to evaluate the symptoms and signs of ADHD;
- **WISC score**
  The Wechsler Intelligence Scale for Children (WISC) [11], developed by David Wechsler, is an individually administered intelligence test for children.

Similarly as in the age prediction task, we attempted to explain the target variable using a linear model, as well as support vector machines and tree-based models, and we benchmarked the prediction accuracy (in terms of MSE) against the sample mean. As in age prediction, we select best parameters for each model class with a grid-search approach. The results obtained in this task (reported in tables 3 and 4) suggest a lack of relationship between such behavioural scores and structural brain information, according to our data.

Table 3: MSEs for SWAN prediction

	MRI		EEG	
	5-CV MSE	Test MSE	5-CV MSE	Test MSE
Sample Mean	–	0.964	-	0.961
Lasso	1.100	0.964	1.123	0.961
XGBoost	1.161	1.032	1.134	1.166
SVR	1.116	0.955	1.104	1.053

Table 4: MSEs for WISC prediction

	MRI		EEG	
	5-CV MSE	Test MSE	5-CV MSE	Test MSE
Sample Mean	–	264.59	–	264.93
Lasso	270.88	255.97	273.81	264.93
XGBoost	291.98	297.06	297.27	267.25
SVR	268.85	245.77	274.96	266.60

7 Prediction with EEG

As an initial direction of the project we explored models for predicting the presence of a given psychiatric disease using EEG data. Accuracy scores of these models show that there seems not to be enough signal in the data to obtain significant results in such tasks. In particular, we report results

for two tasks, the first one being to predict whether a patient is affected or not by ADHD disease, while the second one being to classify patients in healthy and non-healthy. Note that both are binary classification tasks. For both tasks we report the balanced accuracy score, which is a conservative choice considering class imbalance. Again, the EEG missing values are imputed with the median. Best parameters for each model class are selected with a grid-search approach.

Table 5: ADHD vs all classification with EEG data: balanced accuracy scores

	5-CV balanced accuracy (mean)	test balanced accuracy
SVC	0.5201	0.5315
Random Forest Classifier	**0.5335**	0.5315
Extra Trees Classifier	0.5203	0.5413
XGBoost Classifier	0.5184	**0.5615**

Table 6: Healthy vs Ill classification with EEG data: balanced accuracy scores

	5-CV balanced accuracy (mean)	test balanced accuracy
SVC	0.5145	**0.5751**
AdaBoost Classifier	0.5098	0.5471
XGBoost Classifier	**0.5173**	0.4817

8 Conclusion

We showed that, in a pediatric sample, a patient's age can be predicted via structural brain information in the form of MRI and EEG data with reasonable accuracy. Assuming that brain age reflects well the patient's anagraphic age if the patient is healthy, we trained a conditional density estimation model on healthy patients only, and used this model to predict brain age for some classes of patients affected by psychiatric disorders. Interestingly, patients affected by autism exhibit a significantly large gap between their anagraphic age and their predicted brain age; such finding suggests that information about brain age discrepancy may be helpful in diagnosing specific disorders.
A further direction for our study would be to use Factor Analysis (or another latent variable model) on the behavioural data. Finding interpretable factors that show correlation with the age discrepancy (as we define it above) could lead to a better understanding on how this discrepancy might influence the various cognitive areas (e.g. memory, reasoning or attention).

References

[1] O. Al Zoubi, C. Ki Wong, R. T. Kuplicki, H.-w. Yeh, A. Mayeli, H. Refai, M. Paulus, and J. Bodurka, "Predicting age from brain eeg signals—a machine learning approach," *Frontiers in Aging Neuroscience*, vol. 10, p. 184, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2018.00184

[2] M. J. Aminoff, "Chapter 3 - electroencephalography: General principles and clinical applications," in *Aminoff's Electrodiagnosis in Clinical Neurology (Sixth Edition)*, sixth edition ed., M. J. Aminoff, Ed. London: W.B. Saunders, 2012, pp. 37 – 84. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781455703081000030

[3] S. R. Chandra, A. Asheeb, S. Dash, N. Retna, K. V. Ravi Teja, and T. G. Issac, "Role of Electroencephalography in the Diagnosis and Treatment of Neuropsychiatric Border Zone Syndromes," *Indian J Psychol Med*, vol. 39, no. 3, pp. 243–249, 2017.

[4] J. H. Cole, R. Leech, D. J. Sharp, and for the Alzheimer's Disease Neuroimaging Initiative, "Prediction of brain age suggests accelerated atrophy after traumatic brain injury," *Annals of Neurology*, vol. 77, no. 4, pp. 571–581, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.24367

[5] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, pp. 115–124, 2017.

[6] A. Kirsten, S. Linder, and S. Olbrich, "[Perspectives for the Electroencephalogram in Psychiatry]," *Praxis (Bern 1994)*, vol. 107, no. 15, pp. 837–843, Jul 2018.

[7] T. Morita, M. Asada, and E. Naito, "Contribution of Neuroimaging Studies to Understanding Development of Human Cognitive Brain Functions," *Front Hum Neurosci*, vol. 10, p. 464, 2016.

[8] T. J. Polderman, E. M. Derks, J. J. Hudziak, F. C. Verhulst, D. Posthuma, and D. I. Boomsma, "Across the continuum of attention skills: a twin study of the swan adhd rating scale," *Journal of Child Psychology and Psychiatry*, vol. 48, no. 11, pp. 1080–1087, 2007.

[9] J. Rothfuss, F. Ferreira, S. Boehm, S. Walther, M. Ulrich, T. Asfour, and A. Krause, "Noise regularization for conditional density estimation," 2019.

[10] S. M. Smith, D. Vidaurre, F. Alfaro-Almagro, T. E. Nichols, and K. L. Miller, "Estimation of brain age delta from brain imaging," *NeuroImage*, vol. 200, pp. 528 – 539, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811919305026

[11] D. Wechsler, "Wechsler intelligence scale for children; manual." 1949.