Prediction of Fluid Intelligence From T1-Weighted Magnetic Resonance Images

Sebastian Pölsterl, Benjamín Gutiérrez-Becker, Ignacio Sarasua, Abhijit Guha Roy, and Christian Wachinger

Artificial Intelligence in Medical Imaging (AI-Med),
Department of Child and Adolescent Psychiatry,
Ludwig Maximilian Universität, Munich, Germany
{firstname}@ai-med.de

Abstract. We study predicting fluid intelligence of 9-10 year old children from T1-weighted magnetic resonance images. We extract volume and thickness measurements from MRI scans using FreeSurfer and the SRI24 atlas. We empirically compare two predictive models: (i) an ensemble of gradient boosted trees and (ii) a linear ridge regression model. For both, a Bayesian black-box optimizer for finding the best suitable prediction model is used. To systematically analyze feature importance our model, we employ results from game theory in the form of Shapley values. Our model with gradient boosting and FreeSurfer measures ranked third place among 24 submissions to the ABCD Neurocognitive Prediction Challenge. Our results on feature importance could be used to guide future research on the neurobiological mechanisms behind fluid intelligence in children.

1 Introduction

Fluid intelligence [3] is a neuroscientific concept that is closely related to, but distinct from, general intelligence. It refers to the ability to think logically and to solve novel problems. It is believed that fluid reasoning plays a central role in cognitive development from childhood to early adulthood and enables children to acquire other abilities [1]. Anatomically, it is widely believed that higher fluid intelligence is linked to the maturation of the prefrontal cortex [9,25]. Previous neuroscientific findings indicate that brain volumes in parietal, occipital, and temporal as well as frontal cortical areas are related to intelligence [10]. In particular, it has been demonstrated that a region in the anterior prefrontal cortex, known as the rostrolateral prefrontal cortex plays a role in fluid reasoning [25,5]. Moreover, cortical thickness in early childhood has been associated with increased intelligence [14,21].

The ABCD Neurocognitive Prediction Challenge 2019 focuses on studying the relationship between brain and behavioral measures by asking participants to infer fluid intelligence solely from structural T1-weighted magnetic resonance images (MRI). To control for confounding factors, the organizers employed a residualized fluid intelligence score that accounts for a child's brain volume, age

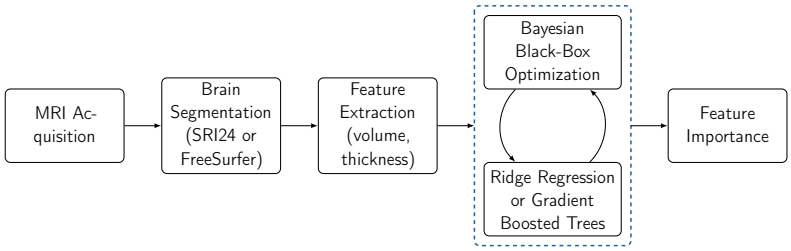

Fig. 1. Overview of our proposed pipeline for the prediction of fluid intelligence from T1-weighted MRI scans using SRI or FreeSurfer features.

at baseline, sex at birth, ethnicity, highest parental education, parental income, parental marital status, and image acquisition site.

For extracting quantitative markers from MRI scans, we use two different approaches in this work. First, we use volumetric measurements provided by the challenge organizers based on the SRI24 atlas [18]. Second, we process MRI scans with FreeSurfer [6] for extracting volume and thickness measurements. In prior work, we obtained competitive results with FreeSurfer features for Alzheimer's prediction [24,2]. Having access to two sets of features describing brain structure allows us to objectively evaluate which set provides more information for predicting fluid intelligence score. Moreover, we compare two approaches for regressing the fluid intelligence score: (i) a linear ridge regression, and (ii) gradient boosted trees. Linear ridge regression is a simple model, which offers the advantage of being less susceptible to overfitting. We use it as a baseline, because linear models are often used in clinical research due to being easy to interpret. Gradient boosted trees can model more complex relationships and are among the best performing prediction methods [26] and have been used in several winning entries to a wide range of Kaggle competitions.[1]

2 Data

We employ data that was provided by The Adolescent Brain Cognitive Development(ABCD) Study [16], which recruited children aged 9-10. The residualized fluid intelligence scores and T1-weighted MRI scans of 3,736 children were provided to participants for training, and 415 samples for validation. Finally, data from 4,402 subjects were provided without fluid intelligence scores for testing. All data was obtained from the National Institute of Mental Health Data Archive.[2]

3 Methods

Fig. 1 illustrates the components of our pipeline for the prediction of fluid intelligence from T1-weighted MRI scans. Scans were acquired according to the

[1] See https://github.com/dmlc/xgboost/blob/master/demo/README.md.
[2] https://nda.nih.gov/edit_collection.html?id=3104

acquisition protocol of the Adolescent Brain Cognitive Development (ABCD) study protocol.[3] For quantifying brain morphometry, we used two different approaches: the SRI atlas and FreeSurfer. For SRI, we used volume measurements of 122 regions of interest extracted by the challenge's organizers based on the SRI24 atlas [18].[4] For FreeSurfer, we first performed automated segmentation of anatomical brain structures on MRI scans using FreeSurfer (version 5.3) [6]. We then extracted 136 volume and thickness measurements, which are all the measures produced by FreeSurfer scripts `asegstats2table` and `aparcstats2table`. These brain measures constituted the input to our machine learning models, linear ridge regression and gradient boosted trees [7,8]. For both, Bayesian black-box optimization for hyper-parameter tuning [22] was applied. Finally, we evaluated feature importance of the final models using the recently proposed SHapley Additive exPlanations (SHAP)[13].

3.1 Data-Preprocessing

We normalized all measurements while accounting for outliers by subtracting the median and dividing by the range between the 5% and 95% percentile. Thus, we reduce the impact of outliers and still obtain approximately centered features with equal scale. Finally, the provided residualized fluid intelligence scores in the training data where standardized to zero mean and unit variance; the same transformation as derived from the training data was applied to features and scores in the validation and test data.

3.2 Models

In following, we describe the two prediction models, where we used FreeSurfer features for the linear prediction, and FreeSurfer and SRI features for gradient boosting.

Linear Model For the linear prediction of residualized fluid intelligence score, we combined features selected by automatic feature selection with features derived from literature. In particular, we included regions of the prefrontal cortex, because of results in previous studies [10,9,25]. For automatic feature selection, we employed univariate feature scoring by estimating the mutual information between each feature and the residualized fluid intelligence score [12]. First, we fitted a linear model based on FreeSurfer features describing the respective regions of interest in the prefrontal cortex. Next, we compared the estimated coefficients with the estimated feature importance by mutual information across all FreeSurfer features. After fusing this information, we ultimately selected eight features (see table 3). Our final model was a ridge regression model [11], where we determined the strength of the ℓ_2 penalty by Bayesian black-box optimization

[3] https://abcdstudy.org/images/Protocol_Imaging_Sequences.pdf

[4] See https://nda.nih.gov/data_structure.html?short_name=btsv01 for a full list of volumes.

(see next section). We fitted the model using the implementation in scikit-learn (version 0.18.2) [15].

Gradient Boosting In addition, we selected stochastic gradient boosting [7,8] for predicting fluid intelligence. Gradient boosting performs functional gradient descent to find a function f that minimizes the squared loss:

$$\arg\min_{f} \quad \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f(\mathbf{x}_i)\right)^2, \tag{1}$$

where y_i denotes the standardized residualized fluid intelligence of the i-th subject and $\boldsymbol{x}_i$ is the feature vector of volume measurements derived from the MRI scan of the i-th subject. Gradient boosting constructs the function f in a greedy stagewise manner by fitting a simple base model g with parameters θ at the m-th iteration to the residuals of f from the previous iteration:

$$\arg\min_{\beta_m,\theta_m} \quad \frac{1}{N}\sum_{i=1}^{N}\Big(y_i - f^{(m-1)}(\mathbf{x}_i) - \beta_m g(\mathbf{x}_i \,|\, \theta_m)\Big)^2, \tag{2}$$

where $\beta_m \in \mathbb{R}$ is a weighting factor. The final estimated ensemble model after M iterations is given by

$$\hat{f}(\mathbf{x}) = \sum_{i=0}^{M} \beta_m g(\mathbf{x}_i \,|\, \theta_m). \tag{3}$$

Here, at each iteration, we employ a regression tree as base model g. To prevent overfitting, we fit the base model to a randomly selected subsample half the size of the whole training data. In addition, we use ℓ_1 regularization on the weights β_m to further reduce overfitting. We used the implementation available in XGBoost [4].

3.3 Hyper-Parameter Selection

The performance of machine learning models depends to a large extent on the choice of hyper-parameters, e.g., the number of base learners in an ensemble. Traditionally, hyper-parameters have been selected by an expert or a grid search is performed. Grid search finds the best performing hyper-parameter configuration by evaluating all possible configurations selected a priori by an expert. However, it is often unknown a priori which values are suitable for a given machine learning task, which makes defining a list of values challenging. In contrast, Bayesian black-box optimization (see e.g. [22]) only requires specifying a prior distribution for each hyper-parameter. Most importantly, it can explore the whole domain of values and not only a list of pre-defined values. It has been shown that Bayesian hyper-parameter optimization outperforms grid search for many machine learning tasks [19].We performed hyper-parameter search using scikit-optimize.[5]

[5] https://scikit-optimize.github.io

Table 1. Priors for hyper-parameters used in Bayesian hyper-parameter optimization. The first four rows refer to parameters from gradient boosting models, the last row is the only parameter optimized for the linear model. We denote a real-valued random variable drawn from a uniform distribution over the domain $[a, b] \subset \mathbb{R}$ as $X \sim \mathcal{U}(a, b)$. A random variable drawn from a log-uniform distribution is denoted as $X \sim \mathcal{LU}(a, b)$, where $X = \log_{10} Y$ with $Y \sim \mathcal{U}(10^a, 10^b)$. Finally, a random variable drawn from integers in the interval $[a, b] \subset \mathbb{Z}$ is denoted as $X \sim \mathcal{U}_{\text{int}}(a, b)$.

Parameter	Prior
Iterations	$\mathcal{U}(10, 1000)$
Max. depth	$\mathcal{U}(1, 5)$
Learning rate	$\mathcal{LU}(10^{-5}, 1.25)$
ℓ_1 regularization	$\mathcal{LU}(10^{-6}, 2^{12})$
ℓ_2 regularization	$\mathcal{LU}(10^{-6}, 2^{12})$

For each proposed hyper-parameter configuration, we estimated the squared error over the validation set and used the sum of median and standard deviation of all errors as optimization criteria. We used Gaussian processes as surrogate models and probabilistically chose one of three acquisition functions at each iteration: (i) expected improvement (EI), (ii) lower confidence bound (LCB), or (iii) probability of improvement (PI). Initially, we assign each acquisition function equal weight $w_i = 0$ ($i = \{1, 2, 3\}$). For the t-th iteration, we independently optimized each acquisition function to propose a candidate hyper-parameter optimization $\mathbf{h}_i^{(t)}$. The final proposal $\mathbf{h}_*^{(t)}$ was selected probabilistically according to the softmax distribution with $p_i = \frac{w_i}{\sum_{k=1}^{3} w_i}$. After retrieving the corresponding prediction error and updating the surrogate model accordingly, we update the weights such that $w_i^{(t+1)} = w_i^{(t)} - \mathbb{E}(\mathbf{h}_*^{(t)})$. This process was repeated for 100 iterations. The best-performing hyper-parameter configuration was used for prediction. Our choice of prior distributions is summarized in table 1.

3.4 Feature Importance

While gradient boosted trees are a potentially powerful model to solve a variety of prediction tasks, their black-box nature is often a barrier for the adoption of such model in the clinic. To be able to interpret complex non-linear machine learning methods, such as gradient boosting trees, we rely on Shapley values, which are a classic solution in game theory to determine the distribution of credits to players participating in a cooperative game [20,23]. In particular, we employ the recently proposed SHAP (SHapley Additive exPlanations) values, which belong to the class of additive feature importance measures [13]. We describe the assignment of importance values to features in detail in [17].

4 Results

We describe the results for the three models: (i) linear model with FreeSurfer-based features, (ii) gradient boosting with SRI-based features, and (iii) gradient

Table 2. Performance on training, validation and test set. MSE: mean squared error. MAE: mean absolute error. GBM: Gradient Boosting Model.

		Linear		GBM (SRI24)		GBM (FreeSurfer)	
	Subjects	MSE	MAE	MSE	MAE	MSE	MAE
Training	3,736	85.492	7.304	61.657	6.241	47.547	5.487
Validation	415	71.277	6.521	71.477	6.579	69.653	6.557
Test	4,402	93.215	—	94.103	—	92.563	—

boosting with FreeSurfer-based features. The performance of all our models for the prediction of residualized fluid intelligence is summarized in table 2. First, we are comparing models based on prediction performance. In the second part, we are inspecting models in more detail via SHAP values.

4.1 Prediction Performance

Our linear model using FreeSurfer features ranked tenth among 24 submissions to the ABCD Neurocognitive Prediction Challenge with a difference of 1.0854 and 0.7179 to the first and second placed team, respectively. Results indicate that predicting residualized fluid intelligence from MRI-derived volume measurements is a challenging task for a linear model. In particular, the proposed model struggles to reliably predict residualized fluid intelligence at the extremes of the distribution, i.e., very low or very high values. Consequently, we observe a relatively high mean squared error, which is an order of magnitude larger than the mean absolute error. Interestingly, we obtained a lower error on the validation data than the training data. However, the error on the test data is relatively high. Although we do not have access to the test data, we believe the main increase in test error can be attributed to mispredictions at the extreme ends of the fluid intelligence distribution. We believe this to be a reasonable assumption since our model is a linear model with only few features, such that overfitting is unlikely to be a problem.

Our gradient boosting model based on SRI Features ranked 17th on the challenge's test data. Although gradient boosting allows for modelling non-linear relationships between MRI-derived features and residualized fluid intelligence, we note that table 2 shows that the prediction error on the training set improved, but as before remains high, which indicates problems in predicting the extremes of the distribution. Moreover, worse performance on the validation and test set indicates that generalization to unseen data seems to be an issue to some extent. Unfortunately, we do not have access to the test data and cannot compute the mean absolute error. Therefore, it is unclear to which extent the relatively high mean squared error on the test set is due to overfitting on the training data or due to larger errors at the extreme ends of the fluid intelligence distribution.

Table 3. Estimated coefficients for all features used in the linear ridge regression model.

Feature	Coefficient
Volume of the brain mask	-0.284
Right Hippocampus volume	0.131
Right Ventral Diencephalon volume	0.206
Left Paracentral thickness	0.082
Right Cortical White Matter volume	0.113
Right Fusiform thickness	-0.050
Right Superior Temporal thickness	-0.009
Left Lateral Orbitofrontal thickness	-0.066

By using gradient boosting with FreeSurfer features, we were able to achieve lower prediction error than with SRI-based features (see table 2). Our model ranked third place among 24 submissions to the ABCD Neurocognitive Prediction Challenge with a difference of 0.0652 and 0.4327 to the second and first placed team, respectively. The model outperformed the linear regression model from above by 1.573 on the validation data and 0.652 on the test data.

4.2 Feature Importance

To get a better understanding of which brain regions drive the predictions, we inspected feature importance for each individual model. Feature importance for the linear model is readily available from the model's coefficients. The estimated coefficients of all features included in the final linear model are summarized in table 3. Since we unified the scale of features, we can directly compare the absolute value of coefficients to obtain a ranking. Our model assigned the highest importance to the volume of the brain mask, where an increase in volume is associated with a decrease in fluid intelligence score. The volume of the right ventral diencephalon is ranked second and positively correlated with fluid intelligence.

For gradient boosting models, we computed SHAP values for each subject and feature in the training data. By ranking features according to the mean absolute SHAP value, we get an overall ranking of features according to their importance. We list the top 20 SRI features by mean absolute SHAP value ϕ in fig. 2a. The top ranked features are the volumes of left parahippocampal gyrus ($\phi = 0.0204$) and pons white matter ($\phi = 0.0202$), while right parahippocampal gyrus ($\phi = 0.0155$) follows with a slightly lower SHAP value. We note that the maximum mean absolute SHAP value of 0.0204 for left parahippocampal gyrus is relatively small, compared the overall scale of the standardized residualized fluid intelligence score ($\min = -4.1905$, $\max = 3.0276$). Therefore, we do not observe a strong association between a single brain region and fluid intelligence. Instead, it seems that the interplay between multiple region seems to be more important.

Fig. 2. Gradient boosting with SRI: (a) Top 20 features sorted by mean absolute SHAP value $\bar{\phi}_j$. (b) SHAP values of top 20 features for each subject in the training data. In each row SHAP values ϕ_j for each subject are plotted horizontally, stacking vertically to avoid overlap. Each dot is colored by the value of that feature, from low (blue) to high (red). If the impact of the feature on the model's prediction varies smoothly as its value changes then this coloring will also appear smooth.

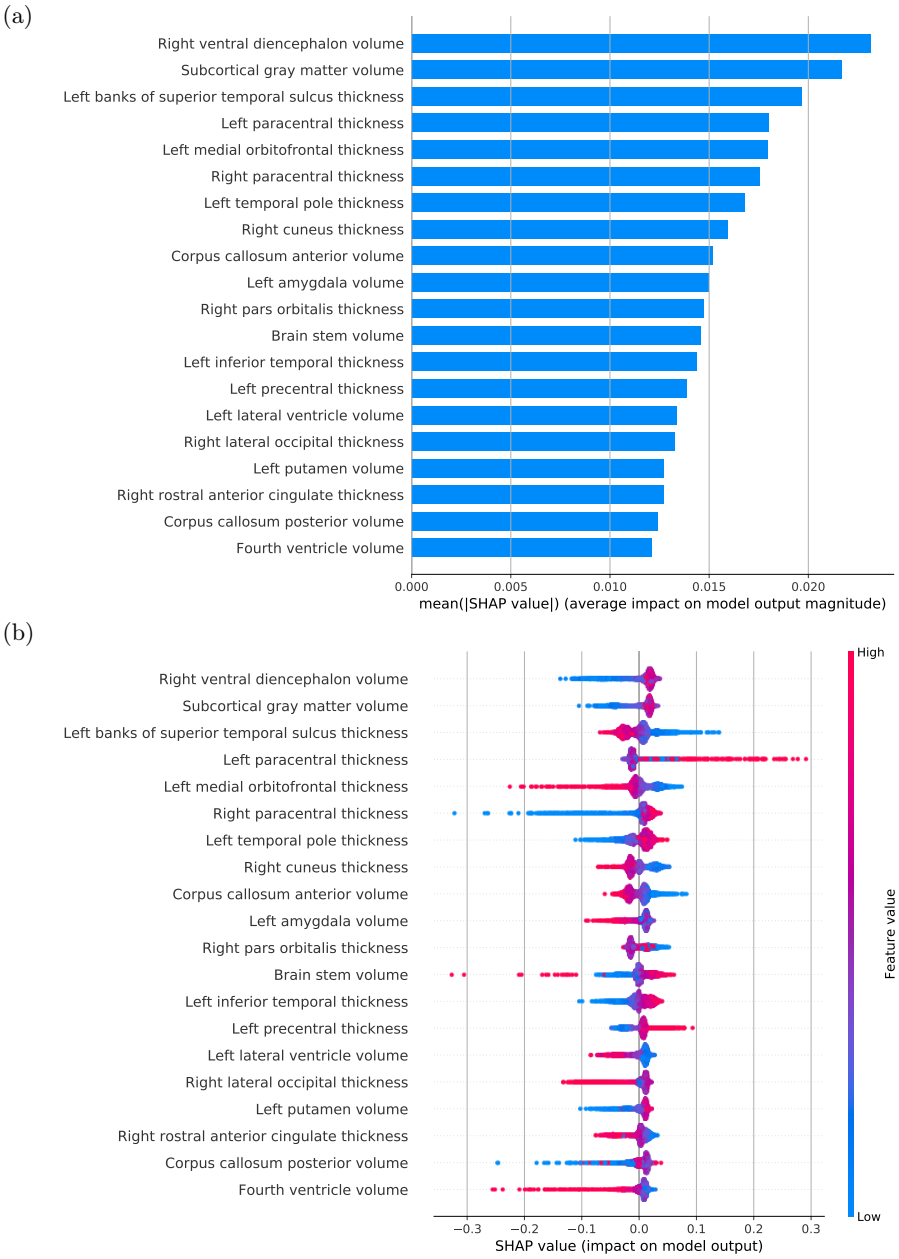

Fig. 3. Gradient boosting with FreeSurfer: (a) Top 20 features sorted by mean absolute SHAP value $\bar{\phi}_j$. (b) SHAP values of top 20 features for each subject in the training data. In each row SHAP values ϕ_j for each subject are plotted horizontally, stacking vertically to avoid overlap. Each dot is colored by the value of that feature, from low (blue) to high (red). If the impact of the feature on the model's prediction varies smoothly as its value changes then this coloring will also appear smooth.

The top 20 FreeSurfer features by mean absolute SHAP value ϕ are listed in fig. 3a. Nine out of the top 20 features represent volume measurements, while the remainder represents thickness measurements. The top ranked feature are the volume of the right ventral diencephalon ($\phi = 0.0231$), followed by subcortical gray matter volume ($\phi = 0.0217$) and left banks of superior temporal sulcus thickness ($\phi = 0.0197$). As we observed above for SRI features, the top ranked feature still has a relatively small effect on the prediction of the model on its own; the right ventral diencephalon volume with $\phi = 0.0231$ on average only contributes a small fraction to the model's predictions. Therefore, single FreeSurfer features, too, do not have a strong association with fluid intelligence.

When comparing all three models, we note that there is little overlap in the most discriminative features. The linear model and gradient boosting with SRI features, both identified hippocampus, fusiform, and superior temporal gyrus to be associated with fluid intelligence. Rankings based on FreeSurfer features used in the linear model and gradient boosting share ventral diencephalon volume and paracentral thickness. Interestingly, while the linear model assigned the highest importance to the overall brain volume, it is absent from the top 20 features when using gradient boosting. Considering that the prediction error for all models is relatively high and that the contribution of individual features are low (small mean SHAP value), it is not surprising that we were unable to identify a clear signature of feature importance across all models considered here.

In addition to the mean absolute SHAP value, we can also look at individual, subject-specific SHAP values depicted in fig. 2b for SRI features and fig. 3b for FreeSurfer features. Figure 2b shows that left and right parahippocampal gyrus volume are negatively associated with fluid intelligence, meaning higher volumes result in a smaller predicted residualized fluid intelligence score. In contrast, the association for pons white matter is positive. Figure 3b for FreeSurfer features shows that higher volume of right ventral diencephalon is associated with an increase in fluid intelligence score, which is consistent with the linear model in tab. 3. In addition, higher thickness of the left banks of superior temporal sulcus are associated with a decrease in fluid intelligence score.

Moreover, fig. 2b indicates that the absolute SHAP value for volume of left inferior temporal gyrus and left hippocampus for a small number of subjects is orders of magnitude larger than that of the average patient. For these subjects, the predicted fluid intelligence score can drop by up to ≈ 0.25 when including one of these features. Thus, these volumes have a very high impact for these patients, although the global impact is modest. Similar effects can be observed in fig. 3b for FreeSurfer features. For instance, we can observe SHAP values up to ≈ -0.3 for brain stem volume, which indicates a significantly bigger importance than the average SHAP value of 0.0146. For these subjects, the brain stem volume highly impacts the model's prediction.

Finally, non-linear effects of estimated models can be observed. Figure 2b shows that for left middle frontal gyrus (orbital part) and right lobule ix of cerebellar hemisphere, large volumes can be associated with either positive or negative SHAP values. Figure 3b shows that brain stem volume seems to be

non-linearly correlated with fluid intelligence: large volumes can be associated with either positive or negative SHAP values.

5 Conclusion

We studied the prediction of fluid intelligence from T1-weighted magnetic resonance images based on features derived from the SRI24 atlas and FreeSurfer. We proposed a linear ridge regression model with 4 volume, and 4 thickness measurements of the brain, and compared it to gradient boosting models using 122 (SRI24) and 136 (FreeSurfer) measurements of the brain, respectively. We experienced predicting fluid intelligence from MRI scans to be a generally difficult task. Our experiments showed that using FreeSurfer features offers a slight advantage in prediction accuracy over SRI24 features, and further gradient boosting over our linear approach. Our gradient boosting model with FreeSurfer ranked third place among 24 submissions to the ABCD Neurocognitive Prediction Challenge. The model indicates that both higher volumes of right ventral diencephalon and higher volumes of subcortical gray matter are associated with an increase in fluid intelligence score. At the same time, we did not find sufficient evidence that fluid intelligence is influenced by only one or a few brain regions. Therefore, we conclude that future research should focus on the interaction between different regions in the brain and their development over time to get a better understanding of which neurobiological factors could drive fluid intelligence.

Acknowledgements This research was partially supported by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B).

References

1. Blair, C.: How similar are fluid cognition and general intelligence? a developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. The Behavioral and brain sciences **29**, 109–25; discussion 125–60 (2006)
2. Bron, E.E., Smits, M., Van Der Flier, W.M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J.M., Steketee, R.M., Orellana, C.M., Meijboom, R., et al.: Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: the caddementia challenge. NeuroImage **111**, 562–579 (2015)
3. Carroll, J.B.: Human cognitive abilities. Cambridge University Press (1993)
4. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794 (2016)
5. Ferrer, E.: Fluid reasoning and the developing brain. Frontiers in Neuroscience **3**(1) (2009)
6. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (2012)
7. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. The Annals of Statistics **29**(5), 1189–1232 (2001)

8. Friedman, J.H.: Stochastic gradient boosting. Computational Statistics & Data Analysis **38**(4), 367–378 (2002)
9. Gray, J.R., Chabris, C.F., Braver, T.S.: Neural mechanisms of general fluid intelligence. Nature Neuroscience **6**(3), 316–322 (2003)
10. Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., Alkire, M.T.: Structural brain variation and general intelligence. NeuroImage **23**(1), 425–433 (2004)
11. Hoerl, A.E., Kennard, R.W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics **12**(1), 55–67 (1970)
12. Kozachenko, L.F., Leonenko, N.N.: Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsii **23**(2), 9–16 (1987)
13. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30. pp. 4765–4774 (2017)
14. Narr, K.L., Woods, R.P., Thompson, P.M., Szeszko, P., Robinson, D., Dimtcheva, T., Gurbani, M., Toga, A.W., Bilder, R.M.: Relationships between IQ and regional cortical gray matter thickness in healthy adults. Cerebral Cortex **17**(9), 2163–2171 (2006)
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
16. Pfefferbaum, A., Kwon, D., Brumback, T., Thompson, W.K., Cummins, K., Tapert, S.F., Brown, S.A., Colrain, I.M., Baker, F.C., Prouty, D., et al.: Altered brain developmental trajectories in adolescents after initiating drinking. American Journal of Psychiatry **175**(4), 370–380 (2018)
17. Pölsterl, S., Gutiérrez-Becker, B., Sarasua, I., Guha Roy, A., Wachinger, C.: An Auto-ML approach for the prediction of fluid intelligence from MRI-derived features. In: Adolescent Brain Cognitive Development Neurocognitive Prediction Challenge (ABCD-NP-Challenge) (2019)
18. Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A.: The SRI24 multichannel atlas of normal adult human brain structure. Human Brain Mapping **31**(5), 798–819 (2010)
19. Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N.: Taking the human out of the loop: A review of bayesian optimization. Proc. of the IEEE **104**(1), 148–175 (2016)
20. Shapley, L.S.: A value for n-person games. Contributions to the Theory of Games **2**(28), 307–317 (1953)
21. Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., Giedd, J.: Intellectual ability and cortical development in children and adolescents. Nature **440**(7084), 676–679 (2006)
22. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian Optimization of Machine Learning Algorithms. In: Advances in Neural Information Processing Systems 25. pp. 2951–2959 (2012)
23. Strumbelj, E., Kononenko, I.: Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems **41**(3), 647–665 (2014)
24. Wachinger, C., Reuter, M., Initiative, A.D.N., et al.: Domain adaptation for alzheimer's disease diagnostics. Neuroimage **139**, 470–479 (2016)
25. Wright, S., Matlen, B., Baym, C., Ferrer, E., Bunge, S.: Neural correlates of fluid reasoning in children and adults. Frontiers in Human Neuroscience **2**, 8 (2008)
26. Zhang, C., Liu, C., Zhang, X., Almpanidis, G.: An up-to-date comparison of state-of-the-art classification algorithms. Expert Systems with Applications **82**, 128–150 (2017)