Improving the reliability of fMRI-based predictions of intelligence via semi-blind machine learning

Gabriele Lohmann*[1,2], Samuel Heczko[2], Lucas Mahler[2], Qi Wang[2], Julius Steiglechner[1,2], Vinod J. Kumar[2], Michelle Roost[3], Jürgen Jost[3,4,5], and Klaus Scheffler[1,2]

[1] *Department of Biomed. Magn. Res. Imaging, University Hospital Tübingen, Germany*
[2] *Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany*
[3] *Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany*
[4] *Center for Scalable Dynamics and Artificial Intelligence, Leipzig University, Germany*
[5] *Santa Fe Institute for the Sciences of Complexity, New Mexico, USA*

Abstract

Predicting neuromarkers for cognitive abilities using fMRI has been a major focus of research in the past few years. However, it has recently been reported that many thousands of participants are required to obtain reproducible results (Marek et al (2022)). This appears to be a major impediment to obtaining neuromarkers from fMRI because large sample sizes are typically not available in neuroimaging studies. Here we show that the out-of-sample prediction accuracy can be dramatically improved by supplementing fMRI with readily available non-imaging information so that reliable predictive modeling becomes feasible even for small sample sizes. Specifically, we introduce a novel machine learning method that predicts intelligence from resting-state fMRI data, leveraging educational level as supplementary information. We refer to our approach as "semi-blind machine learning (SML)" because it operates under the assumption that supplementary information, such as educational level, is available for subjects in both the training and test sets. This setup closely mirrors real-world scenarios, especially in clinical contexts, where patient background information typically exists and can be utilized to boost prediction accuracy. However, guarding against bias is crucial. Subjects should not be categorized as more intelligent simply based on their higher education levels. Therefore, our approach contains a component explicitly designed for bias control. We have applied our method to three different data collections and observed marked improvements in prediction accuracies across a wide range of sample sizes. We anticipate that semi-blind machine learning provides a promising approach to fMRI-based predictive modelling with the potential for a wide range of future applications.

Introduction

The prediction of individual behaviour from the fMRI-based functional connectome has been investigated in a number of studies, e.g. [1–13], for a review see [14].

*Corresponding author: lohmann@tuebingen.mpg.de

However, the viability of this line of research has sparked extensive discussions. Major problems already became apparent at the ADHD-200 Global Competition of 2011 in which researchers were tasked with developing machine learning algorithms for diagnosing ADHD from rsfMRI data [15]. Surprisingly, the team that scored best in this competition did not use any imaging data at all, but instead relied entirely on supplementary non-imaging information [16], see also [17]. And more recently, Marek et al. [18] reported that many thousands of participants are needed to obtain reliable predictions of cognitive abilities using fMRI-based functional connectomes, and even then the prediction accuracies were in fact quite poor. For a debate on this topic see [19–27].

Furthermore, Marek et al. have noted that non-imaging information such as behavioral performance may confound links between brain function and cognitive ability. In fact - much like the ADHD-200 Global Competition - they reported that predictions based on non-imaging information may be more accurate than predictions based on rsfMRI [18, Extended Data Fig 3].

Here we propose to capitalize on such non-imaging information instead of regarding it merely as a confound. Specifically, we introduce a novel machine learning algorithm called "Semi-Blind Machine Learning (SML)" that integrates supplementary non-imaging information into the predictive modelling of the rsfMRI data while carefully controlling potential confounding effects. In essence, our objective is to leverage non-imaging information for improving prediction accuracy while avoiding any potential bias that might be incurred by this process. In SML, the leveraging information is assumed to be known for subjects in both the training and the test set, while the target variable is known only for subjects in the training set. Note that with this design we prioritize prediction accuracy over generalizability. In this work, we will primarily use intelligence as the target variable to be predicted, and educational level as the leveraging factor, but we will also investigate several other target variables.

A straightforward approach of incorporating supplementary information might be to simply average predictions obtained from education levels and neuroimaging data independently. However, as shown in Figure 1, this approach would lead to a strong bias so that subjects with higher education levels would automatically be deemed to be more intelligent. Our goal here is to leverage non-imaging information to improve prediction accuracies while avoiding any such bias. Note that bias in machine learning is a growing concern [28]. In the present context, this issue is particularly relevant because the incorporation of supplementary information would inevitably yield a bias unless countermeasures are taken.

Our proposed approach is based on partial least squares regression combined with ensemble learning. The reason for this algorithmic design is that it permits a straightforward means for bias control. We will show a quantitative comparison of our method against ridge regression which is currently one of the most widely used predictive modeling technique in fMRI. Notably, we have excluded neural network approaches from our consideration. This decision is based on recent findings which demonstrate that neural networks do not yield significantly superior performance in this context when compared to conventional regression techniques [29]. Also, our proposed bias control mechanism is not directly transferable to neural network architectures so that a one-to-one comparison would be difficult.

Our method has some similarities with meta-matching [30] in that it also allows predictions with

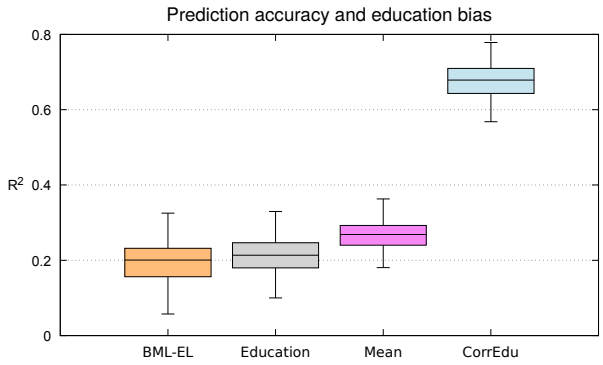

Figure 1: **Education bias in straightforward averaging of predictions (HCP).** The y-axis shows prediction accuracies for intelligence ("CogTotalComp_AgeAdj") using either just rsfMRI data without additional information about education levels ('BML-EL'), or just using education levels ('Education'), or the average of both predictions ('mean'). The rightmost column shows the variance explained by education of the averaged result. Note that this value greatly exceeds the value shown in the 'Education' column. In other words, the averaged prediction is heavily biased so that subjects with higher education levels are automatically deemed to be more intelligent. The rsfMRI data were parcellated using the ICA decomposition into 100 components provided by HCP. Samples sizes for training/test sets were 290 and 100, respectively. 100 random selections of training/test set pairs were used.

small sample sizes. Meta-matching is based on a transfer learning framework in which a large-scale model is adapted to a separate, smaller sample. A significant improvement in prediction accuracy can be achieved although the large scale model may be trained on a completely different set of phenotypes. This improvement results from inherent correlations among phenotypes, see also [31] for more on phenotypic correlations. In our work, we also harness these phenotypic correlations. However, unlike meta-matching, our enhanced prediction accuracy stems from a novel form of feature selection, rather than from the adaptation of a large-scale model.

Semi-blind machine learning (SML)

Predictive modelling generally proceeds in three stages [32]. First, brain parcellations are performed to reduce dimensionality. Second, interactions between the brain parcels or components are estimated. Finally, a classifier or regressor is trained to predict behavioural traits or other quantities of interest. The various methods differ with respect to the choice of the strategies used in those three stages, but some form of network modelling is common to all [33, 34]. The methodological challenges and best practices are discussed in [6, 35–40]. In the following, we introduce our Semi-blind Machine Learning method (SML) for predicting intelligence from rsfMRI data.

Selecting training and test sets. The first step is to select training and test sets. In our experiments, we randomly selected 100 non-overlapping training/test set pairs of various sizes from a given data collection.

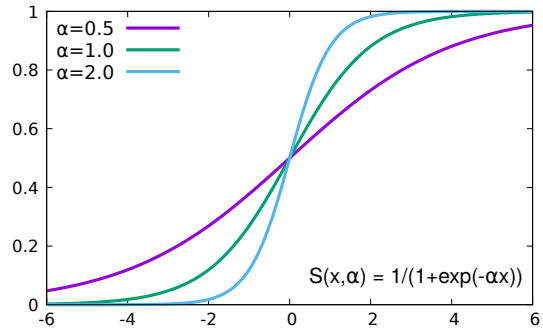

Figure 2: **Illustration of bias control.** The parameter α of the sigmoid function controls the level of randomness in feature selection. Small values of α cause features with negative weights to be more likely to be selected into an ensemble so that the level of randomness in the feature selection process is increased, and the impact of education on the final result is reduced.

Brain parcellations and the connectome. To reduce dimensionality, the fMRI data are parcellated into a set of k regions or spatial components. In the experiments reported below, we used several different parcellation schemes including data driven as well as atlas-based methods. For each subject, a connectivity matrix ("connectome") is computed. Each element of the connectome represents the Pearson linear correlations between fMRI time courses averaged within the parcels or components. Because connectivity matrices are symmetric, we only need to consider the lower triangular parts containing $n = (k \times (k-1))/2$ edges. As proposed by Varoquaux et al. [41], the parcellated connectomes are embedded into the tangent space. In the following, the term 'connectome feature' will be used to denote the edge weight that is computed as the tangent correlation between connectome nodes.

Semi-blind ensemble learning (SML-EL). We have previously proposed a multivariate ensemble learning method to train a prediction model [42]. Here we present an updated version of this method that leverages supplementary non-imaging information – for ease of notation denoted as "L-info". In the following, we assume that the L-info is available for subjects in both the training and the test set.

We use an ensemble learning approach in which multiple learners are trained independently using different subsets of the feature space [43,44]. More precisely, a large number of subsets (ensembles) of connectome features are randomly selected, and are then used in a regression model to predict intelligence based on information contained in the training set. This model is subsequently applied to the test set yielding a prediction of intelligence for each subject in the test set. Finally, the predictions from all ensembles are averaged. Here we use partial least squares regression (PLS) because it is particularly well suited to handle problems where the predictors are highly collinear and where the number of independent variables greatly exceeds the number of data points [45–47].

Feature selection. In our previous work [42], the connectome features that were recruited for an ensemble were selected randomly. Here we propose to use the L-info to guide the selection so that features that are more likely to be predictive of intelligence are more likely to be included in an ensemble. We call this approach "Semi-blind ensemble learning (SML-EL)". The term "Blind

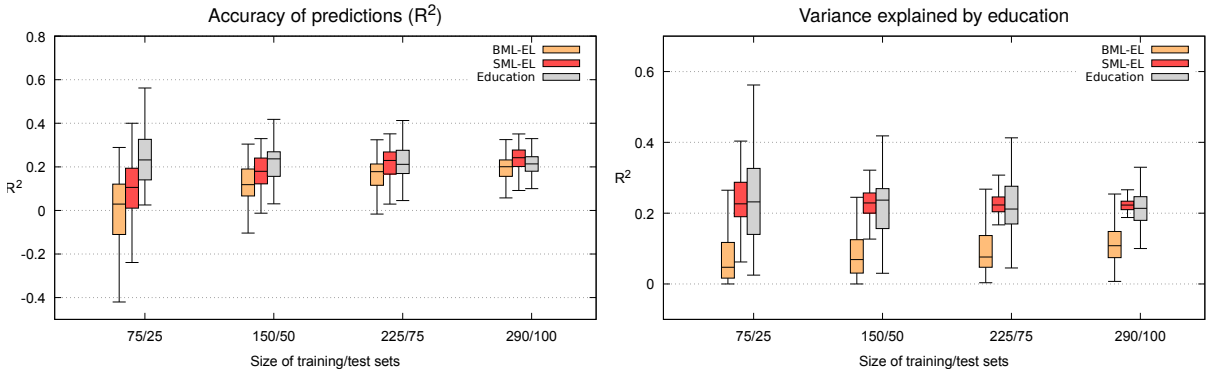

Figure 3: **Prediction accuracies using various sample sizes (HCP).** Samples sizes for training/test sets are shown along the x-axis. A precomputed ICA-based parcellation using 100 components provided by HCP was used ("netmats2.txt"). The left plot shows prediction accuracies, the right plot shows the variance explained by education. If this value is very different from the variance explained in the observed intelligence, it indicates the presence of bias. Note that the bias control mechanism in SML-EL has prevented such a bias. Results using either blind ('BML-EL') or semi-blind ensemble learning ('SML-EL') are shown in orange and red. The predictions using education level alone ('Education') are shown in gray.

ensemble learning (BML-EL)" will be used to denote the case when no L-info is used.

In SML-EL, we conduct a univariate analysis on each connectome feature to determine its relevance for predicting IQ. Specifically, for each connectome feature $i = 1, \ldots, n$, we calculate the linear correlation $\rho_i(train, iq)$ between the IQ scores and the corresponding connectome features in the training set. Since IQ scores are unknown in the test set, education is used as a surrogate, and $\rho_i(test, edu)$ denotes the linear correlation between the education levels and the corresponding connectome features in the test set.

The product of these correlations indicates which connectome features are likely to generalize from the training to the test set:

$$W_i = \rho_i(train, iq) \times \rho_i(test, edu), i = 1, \ldots, n \tag{1}$$

Note that W_i is positive if $\rho_i(train, iq)$ and $\rho_i(test, edu)$ have equal signs, and large positive values of W_i indicate a good correspondence between training and test set, provided education is a useful surrogate for intelligence.

In cases where the training set is much larger than the test set, some samples from the training set are randomly selected and transferred to a size-adjusted test set so that the training set and the adjusted test set become equal in size. Note that this is only done for the purpose of feature selection, and a "leakage" from training to test set does not occur. The subsequent feature selection procedure uses the size-adjusted test set.

The weights W_i are now used for a probabilistic feature selection procedure such that features with large weights have a higher probability of being selected into an ensemble. For this, the weights W_i need to be transformed into non-negative weights P_i. This transformation is a two-step procedure:

first, normalizing the weights through z-scoring, followed by applying a sigmoid function to the normalized values. The sigmoid is defined as

$$S(x, \alpha) = \frac{1}{1 + e^{-\alpha x}} \in [0, 1], \ \alpha > 0 \tag{2}$$

We then use the algorithm by Walker [48] as implemented in [49] for selecting features. More precisely, given features $i = 1, \ldots, n$ with weights $P_i = S(W_i, \alpha)$, this algorithm produces random indices that are consistent with the given weights so that features with high values of P_i are more likely to be selected.

Bias control. The parameter α of the sigmoid function is used as a handle for bias control (Figure 2). More precisely, α is estimated such that education has a similar correlation with the observed as with the predicted intelligence.

For this, we compute an initial prediction of intelligence using SML-EL with $\alpha = 1$ as a starting value. We then compute the correlation of education with intelligence scores predicted by SML-EL for the test set, and compare it to the correlation of education with intelligence scores that were observed in the training set. If the correlation in the test set is larger than the correlation in the training set, the parameter α is reduced by multiplying it by 0.9, otherwise it is increased by multiplying it by 1.1. This procedure is repeated until convergence, i.e. until the two correlations are sufficiently close. In the experiments reported below, iterations stopped when the two correlations did not differ by more than 0.02.

Semi-blind ridge regression (SML-RR). In addition to ensemble learning, we also implemented ridge regression because it is widely used for predictive modelling and it was shown to have a comparable performance to neural network methods in this and similar tasks [11,29,32]. We used the Scikit-learn package [50, 51] for this purpose. Similar to [52], the regularisation parameter was determined using random search using a three-fold nested-loop crossvalidation.

As in SML-EL, we first conduct a univariate analysis on each connectome feature to determine its relevance for predicting IQ using equation 1. But unlike ensemble learning where a probabilistic feature selection approach was used, we implemented a hard threshold for feature selection. Specifically, the connectome features were ranked using the weights of equation 1, and the top fifty percent were used in the subsequent ridge regression.

Experimental results

We applied the proposed methods to three different data collections: HCP [53], AOMIC-ID1000 [54] and ABIDE-1 [55]. The HCP data were used for developing the SML-EL algorithm and for fine-tuning hyperparameters. Note that this may have lead to an algorithmic overfit in the results obtained using HCP.

SML-EL has three hyperparameters, namely the number of ensembles, the size of the ensembles, and the number of PLS components. Using HCP, we found that 1000 ensembles sufficed to reach

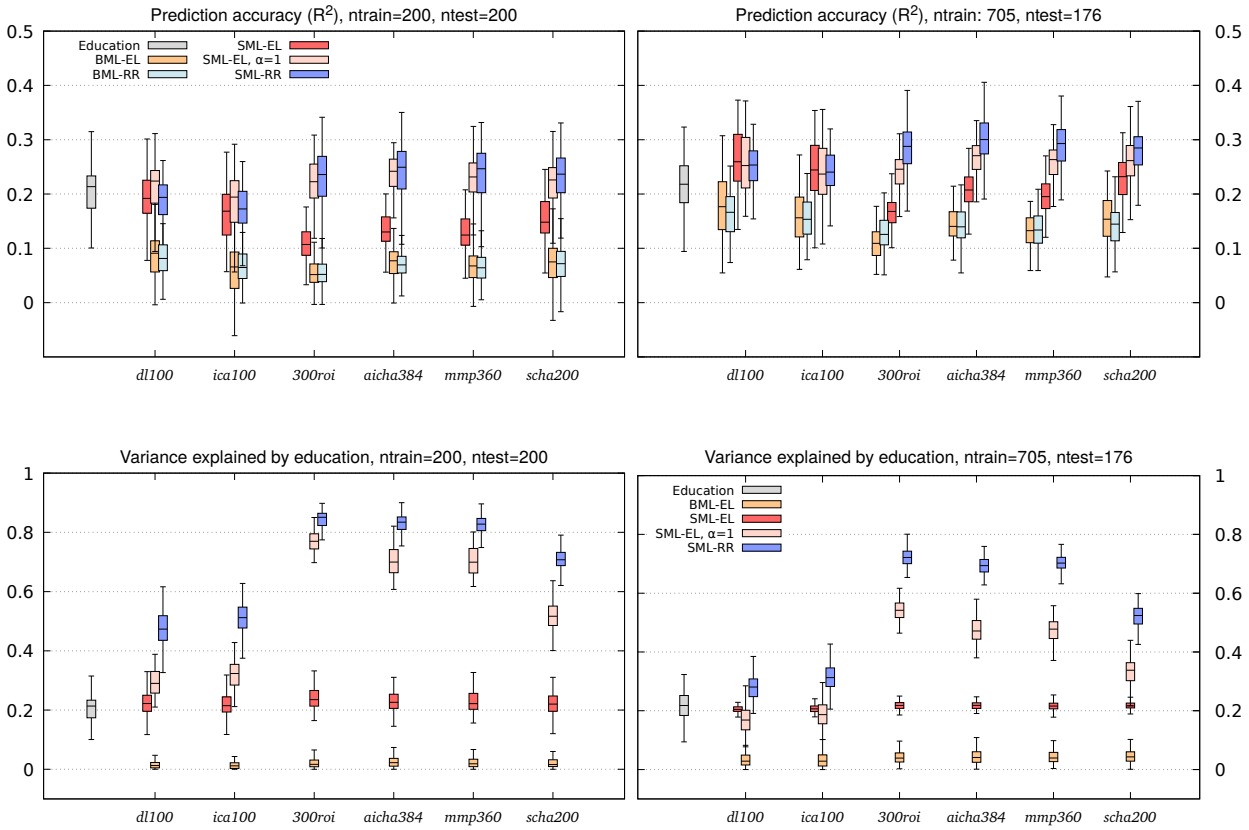

Figure 4: **Comparing SML-EL vs SML-RR using various parcellations (AOMIC).** The sizes for the training/test sets were 200/200 (left) and 705/176 (right), and 100 random selections of training/test set pairs were used. The top plot shows prediction accuracies, the bottom plot shows the variance explained by education. The number of parcels in each parcellation scheme is indicated by its label, for more details see the main text. Both semi-blind methods SML-RR and SML-EL show a marked improvement in prediction accuracy compared to the blind methods BML-EL and BML-RR. Note that the prediction accuracy is higher in SML-RR compared to SML-EL for some parcellation schemes. However, this comes at the price of a very strong education bias, which is not present in SML-EL. For comparison, the results of SML-EL are also shown without bias control ($\alpha = 1$). Note that in this case, the prediction accuracy is approximately the same as for SML-RR, but with considerably less education bias.

convergence. The ensemble sizes were set to 1000, and the number of PLS-components were set to 10. We used these three hyperparameter settings in all experiments reported below.

The accuracy of the predictions is reported as the coefficient of determination between the observed and the predicted intelligence defined as

$$R^2 = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{3}$$

where SSR represents the sum of squared residuals and SST represents the total sum of squares. In the supplementary material, we also report the results using Pearson linear correlations.

In addition, to estimate potential bias, the variance explained by education is reported. It is

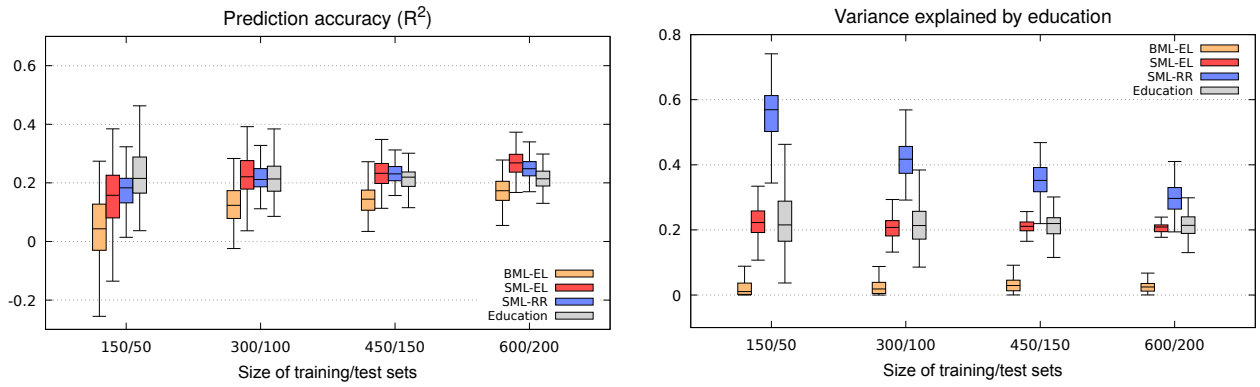

Figure 5: **Accuracy of prediction using various sample sizes (AOMIC).** Samples sizes for training/test sets are shown along the x-axis. 100 random selections of training/test set pairs were used. The data were parcellated into 100 parcels using dictionary learning [56]. The left plot shows prediction accuracies, the right plot shows the variance explained by education. Note that SML-RR is strongly biased by education, whereas SML-EL does not show a bias. Also note that BML-EL shows a clear negative bias, i.e. the correlation of education with the predictions of IQ in BML-EL is much lower than the correlation of education with the observed IQ.

computed by fitting a linear regression model using education as the independent variable and the predicted intelligence scores as the dependent variable. The variance explained by education is then defined as in equation 3 with SST the variance of the predicted intelligence scores and SSR the variance of education levels.

Human Connectome Project (HCP) [53]. We downloaded minimally preprocessed rsfMRI data of 390 unrelated subjects (202 female, 188 male, aged 22-36, median age 28) acquired at 3 Tesla by the Human Connectome Project (HCP), WU-Minn Consortium (`https://db.humanconnectome.org`). For dimensionality reduction we used an ICA parcellation into 100 components and the corresponding connectivity matrices containing partial correlations between the components provided by HCP ("netmats2.txt"). The target variable for predictive modelling was the intelligence score provided by HCP called "CogTotalComp_AgeAdj". As L-info we used information about education levels called "SSAGA_Educ". See the supplementary material for more details on these scores and on the scanning and preprocessing parameters of the fMRI data. The results of SML-EL applied to HCP are shown in Figure 3, see also Supplementary Figures S2,S3.

The Amsterdam Open MRI Collection (AOMIC-ID1000) [54]. We downloaded minimally preprocessed fMRI data of the AOMIC-ID1000 data collection consisting of 928 healthy subjects (`https://openneuro.org/datasets/ds003097`). For 47 of those, fMRI data were not available on the AOMIC website so that only 881 could be included in our study. Subjects were aged between 19 and 26 yrs (mean 22.8) with 457 females and 424 males.

The target variable for predictive modelling was the intelligence score ('iq') as provided by AOMIC. It is based on the Intelligence Structure Test (IST) which measures crystallized and fluid intelli-

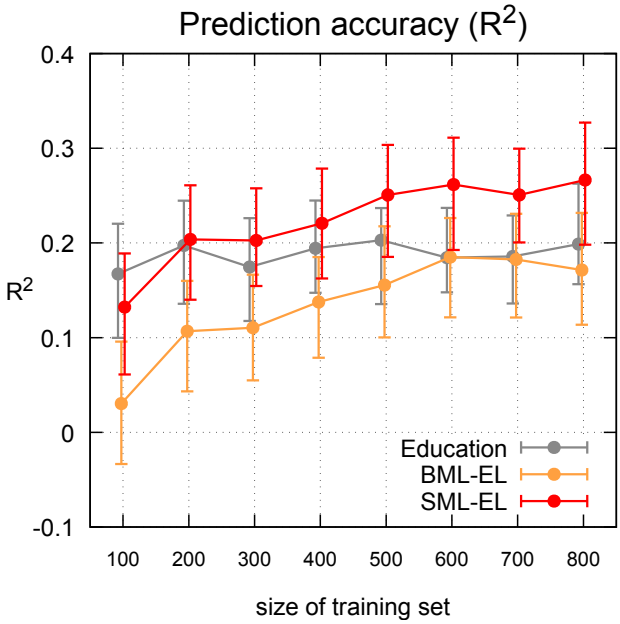

Figure 6: **Prediction accuracy reaches a plateau (AOMIC).** The size of the training sets are shown on the x-axis. The size of the test sets was 80 throughout. The error bars show the upper and lower quartiles. Note that the prediction accuracy initially correlates with the size of the training sets, but reaches a plateau at a sample size of about 600 beyond which it remains relatively constant. Also note that the semi-blind approach (SML-EL) surpasses the prediction accuracy of "Education" for training sample sizes larger than 300, whereas the blind approach (BML-EL) fails in this regard. As before, the data were parcellated into 100 parcels using dictionary learning [56], see also Supplementary Figures S6,S7.

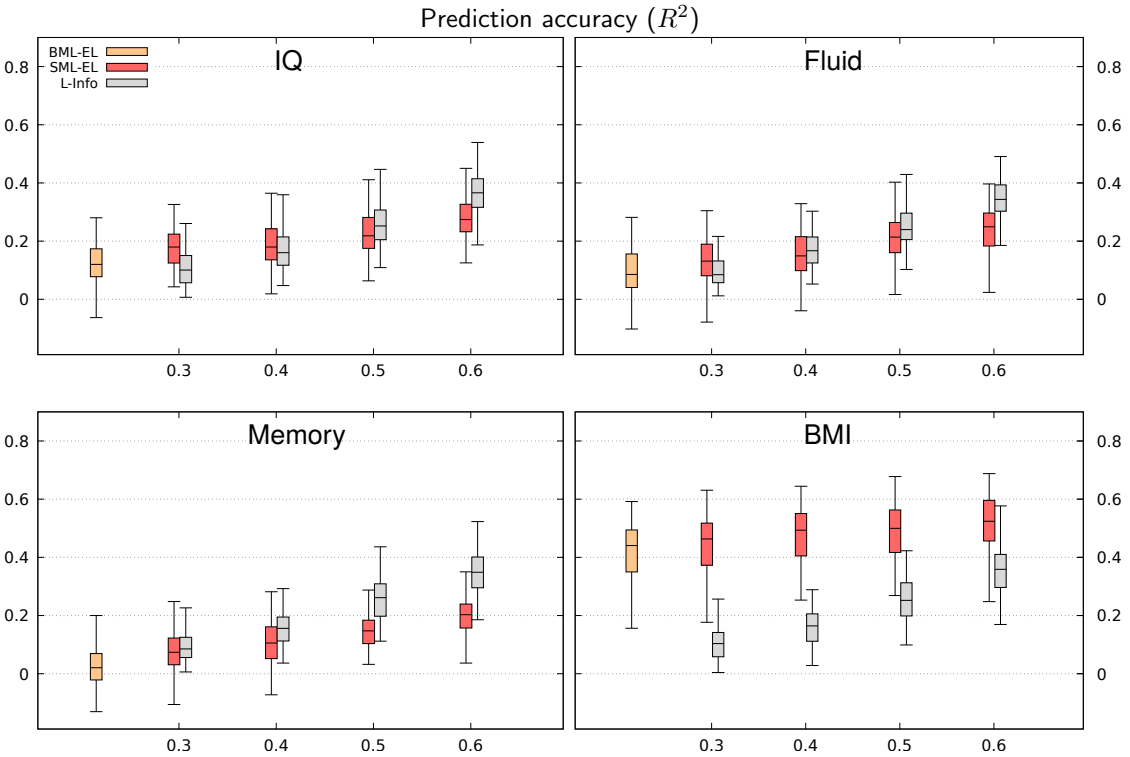

Figure 7: **Predictions of various target variables using simulated leverage information (R^2) - AOMIC.** This figure shows predictions of IQ, fluid intelligence, memory and body mass index (BMI) of the AOMIC collection. The leverage information was randomly generated to various degrees of linear correlation with the target variable as indicated by the numbers on the x-axis ('0.3','0.4','0.5','0.6'). The sizes of training/test sets were 300/100. Note that even weakly correlated L-info helps to improve prediction accuracy. Using strongly correlated L-info alone (without fMRI) may yield higher prediction accuracies than SML-EL, but the results would of course be heavily biased.

gence, as well as memory [57]. We also investigated memory, fluid intelligence, and Body Mass Index (BMI) as further target variables. As auxiliary information we used information about the highest completed educational level ('edu') as provided by AOMIC. Here only two different values were recorded corresponding to whether or not participants had completed college. Of the 881 participants, 411 had completed college, all others had not, see [54] for further details.

Figure 4 shows a comparison of semi-blind ensemble learning and ridge regression results using various brain parcellations. The following parcellations were used [58]: dl100(dictionary learning) [56], ica100, [41], 300roi [59], aicha384 [60], mmp360 [61], scha200 [62]. The number of parcels in each parcellation is indicated by their names. Supplementary Figure S4 shows the corresponding results using Pearson linear correlations between observed and predicted intelligence.

Figure 5 shows results with various sample sizes, see also Supplementary Figure S5. Here, dictionary learning was used for parcellations. Note that even small sample sizes show a clear improvement of prediction accuracy compared to the blind method. Again, SML-RR shows a strong bias, which is not present in SML-EL.

Figure 6 was included to show the effect of increasing training set sizes. Note that the prediction accuracy initially correlates with increases in sample size but reaches a plateau at about 600, indicating that larger sample sizes may not be sufficient to improve prediction accuracies, see also Supplementary Figures S6,S7.

In order to further investigate the role of the L-info for prediction accuracy, we also used simulated L-info and applied it to four different target variables, namely IQ, fluid intelligence, memory and Body Mass Index (BMI) as provided by AOMIC. The simulated L-info was randomly generated with several pre-defined levels of correlation with the target variable ranging from 0.3 to 0.6. The results are shown in Figure 7 and Supplementary Figures S8,S9. Note that even a weak correlation of the simulated L-info with the target variable helps to improve prediction accuracy.

Autism Brain Imaging Data (ABIDE-1) [63]. The ABIDE-1 data collection is typically used to investigate autism spectrum disorders (`https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html`). For the present study we only included data of 442 healthy controls aged 6 to 56 yrs (mean 16.4), while patient data were excluded, see the supplementary material for more details. The reason why we included this specific collection is its greater heterogeneity in comparison to the other two collections. This increased heterogeneity is due to a much wider age range of participants and the data being collected from 16 distinct sites. In this scenario, model prediction becomes significantly more challenging, see also [11]. The target variable for predictive modeling was the FIQ Standard Score (full IQ) as provided by ABIDE-1, and also subject age. Unfortunately, ABIDE-1 does not contain information about education levels. Therefore, we randomly generated data to simulate supplementary non-imaging information with a pre-determined correlation with FIQ and age. Brain parcellations were done using the Dosenbach-160 Atlas [64]. The results of SML-EL applied to ABIDE-1 is shown in Figure 8, see also Supplementary Figures S10,S11. As for the other two data collections, the L-info provided increases in prediction accuracy compared to the blind versions of the algorithm. This was true for both target variables, even though the prediction accuracy for age is markedly better than for FIQ.

Discussion

We have introduced a semi-blind machine learning approach designed to predict intelligence from fMRI data. By incorporating education as leveraging information, we achieved a significant improvement in out-of-sample prediction accuracy. Moreover, we found that sample sizes as small as a few hundred were sufficient to attain prediction accuracies that would typically require thousands of participants. This implies a promising avenue for addressing the concerns raised by Marek et al [18].

As shown in Figure 6, relying solely on increasing sample sizes may not be sufficient for resolving the issue of inadequate prediction accuracy. Therefore, we argue that completely new approaches like the one presented in this paper are needed. Here we have proposed to leverage non-imaging supplementary information for boosting prediction accuracy. We have introduced a bias control mechanism that guarantees that the resulting predictions maintain a correlation with the non-

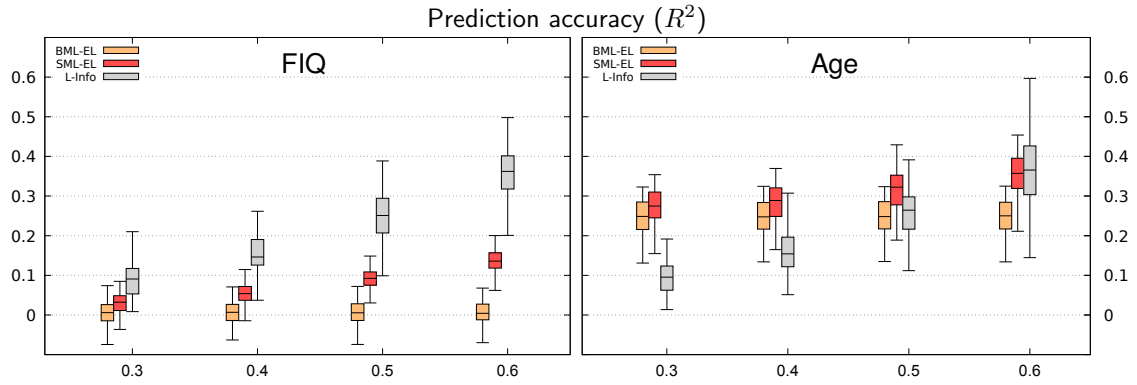

Figure 8: **Predictions of various target variables using simulated leverage information (R^2) - ABIDE-1.** This figure shows predictions of FIQ and age of the ABIDE-1 collection. The leverage information was randomly generated to various degrees of linear correlation with the target variable as indicated by the numbers on the x-axis ('0.3','0.4','0.5','0.6'). The sizes of training/test sets were 300/100.

imaging information similar to that observed in the target variable within the training dataset. Without such a mechanism, the resulting predictions would be heavily biased.

We are aware that our semi-blind models are not generalizable to individuals outside of the given test set because they leverage information from the test set. But we argue it is uncertain whether any model can genuinely claim to be universally applicable to the entire human population so that generalisability may not be a viable objective in any case. Furthermore – because computation times are on the order of a few minutes – semi-blind models can be easily updated if a new set of test subjects arrives.

As noted in the introduction, our approach is related to meta-matching [30] in that it also aims for reducing the sample sizes needed for reliable predictions. But it differs in that it does not require a large-scale model trained on thousands of subjects. Another key distinction is our explicit bias control mechanism which is not included in meta-matching.

We have proposed two different semi-blind machine learning methods both showing a marked improvement of prediction accuracy. The advantage of ensemble learning (SML-EL) combined with a probabilistic feature selection is that it offers a straightforward way of controlling the bias. Ridge regression on the other hand (SML-RR) was found to be considerably more vulnerable to bias.

As expected, we found that information about educational levels is highly predictive of intelligence and clearly outperforms conventional fMRI-based regression methods. However, our proposed semi-blind approach was able to outperform predictions made solely based on education, particularly when using parcellations derived from dictionary learning or ICA. Furthermore, it offers the advantage of avoiding bias, resulting in predictions that reflect intelligence rather than merely educational levels. Additionally, it provides neuroimaging data that can be used to identify brain regions associated with cognitive abilities.

We found that the parcellation schemes with fewer parcels (*ica100, dl100*) produced better results

than more fine-grained parcellations (*mmp360, scha200, aicha384*). This is likely due to the fact that a semi-blind feature selection is more challenging if more features are available. Another explanation might be that *ica100* and *dl100* are both data-driven and are therefore optimized for the data collection at hand. However, in our experiments we have used small subsets of data while these two parcellations were in fact based on much larger assemblies.

In summary, we conclude that semi-blind machine learning provides a promising approach to fMRI-based predictive modelling that lends itself to a wide range of potential future applications.

Software

The software will be made public soon.

Acknowledgments

This work was partially funded by DFG-SPP 2041 "Computational Connectomics".
Samuel Heczko was supported by the International Max Planck Research School for the Mechanisms of Mental Function and Dysfunction (IMPRS-MMFD) .
Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

[1] FX Castellanos, A Di Martino, RC Craddock, AD Mehta, and MP Milham. Clinical applications of the functional connectome. *Neuroimage*, 80:527–540, 2013. doi:10.1016/j.neuroimage.2013.04.083.

[2] S. Smith, T. Nichols, D. Vidaurre, A. Winkler, T.E.J. Behrens, M.F. Glasser, K. Ugurbil, D.M. Barch, D. Van Essen, and K. Miller. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci*, 18:1565–1567, 2015. https://doi.org/10.1038/nn.4125.

[3] E.S. Finn, X. Shen, D. Scheinost, M.D. Rosenberg, J. Huang, M.M. Chun, X. Papademetris, and R.T Constable. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18:1664–1671, 2015.

[4] X. Shen, E.S. Finn, D. Scheinost, M.D. Rosenberg, M.M. Chun, X. Papademetris, and R.T. Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc*, 12(3):506–518, 2017. doi:10.1038/nprot.2016.178.

[5] R.E. Beaty, Y.N. Kenett, A.P. Christensen, M.D. Rosenberg, M. Benedek, Q. Chen, A. Fink, J. Qiu, T.R. Kwapil, M.J. Kane, and P.J. Silvia. Robust prediction of individual creative ability from brain functional connectivity. *PNAS*, 115:1087–1092, 2018.

[6] J. Dubois, P. Galdi, L.K. Paul, and R. Adolphs. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Phil. Trans. R. Soc. B*, 373(20170284), 2018. http://dx.doi.org/10.1098/rstb.2017.0284.

[7] A.A. Joshi, J. Li, H. Akrami, and R.M. Leahy. Predicting cognitive scores from resting fMRI data and geometric features of the brain. In Elsa D. Angelini and Bennett A. Landman, editors, *Medical Imaging 2019: Image Processing*, volume 10949, pages 619 – 625. International Society for Optics and Photonics, SPIE, 2019.

[8] B. Heinrichs and S. Eickhoff. Your evidence? Machine learning algorithms for medical diagnosis and prediction. *Human Brain Mapping*, 41:1435–1444, 2020. doi:10.1002/hbm.24886.

[9] C. Sripada, M. Angstadt, S. Rutherford, A. Taxali, and K Shedden. Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping*, 41(12):31863197, 2020. https://doi.org/10.1002/hbm.25007.

[10] J. Chen, A. Tam, V. Kebets, C. Orban, L.Q.R. Ooi, S. Marek, N. Dosenbach, S. Eickhoff, D. Bzdok, A.J. Holmes, and et al. Shared and unique brain network features predict cognitive, personality, and mental health scores in the abcd study. *Nat Comm*, 13(2217), 2022. https://doi.org/10.1038/s41467-022-29766-8.

[11] N. Traut, K. Heuer, G. Lemaitre, and et al. Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery. *Neuroimage*, 255, 2022. https://doi.org/10.1016/j.neuroimage.2022.119171.

[12] S. Rutherford, S.M. Kia, T. Wolfers, and et al. The normative modeling framework for computational psychiatry. *Nat Protoc*, 17:17111734, 2022. https://doi.org/10.1038/s41596-022-00696-5.

[13] P. McClure, D. Moraczewski, K.C. Lam, A. Thomas, and F. Pereira. Improving the Interpretability of fMRI Decoding Using Deep Neural Networks and Adversarial Robustness. *Aperture Neuro*, pages 1–17, 2023. https://doi.org/10.52294/001c.85074.

[14] B.H. Vieira, G.S.P. Pamplona, K. Fachinello, A.K. Silva, M.P. Foss, and C.E.G. Salmon. On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence*, 93(101654), 2022. https://doi.org/10.1016/j.intell.2022.101654.

[15] http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html, 2011.

[16] M.R.G. Brown, G.S. Sidhu, R. Greiner, N. Asgarian, M. Bastani, P.H. Silverstone, A.J. Greenshaw, and S.M. Dursun. ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Front. Syst. Neurosci*, 6, 2012. https://doi.org/10.3389/fnsys.2012.00069.

[17] A. Omidvarnia, L. Sasse, D.I. Larabi, F. Raimondo, F. Hoffstaedter, J. Kasper, J. Dukart, M. Petersen, B. Cheng, G. Thomalla, S.B. Eickhoff, and K.R. Patil. Is resting state fMRI better than individual characteristics at predicting cognition? *bioRxiv*, 2023. https://doi.org/10.1101/2023.02.18.529076.

[18] S. Marek, B. Tervo-Clemmens, F.J. Calabro, and et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603:654–660, 2022. https://doi.org/10.1038/s41586-022-04492-9.

[19] T. Spisak, U. Bingel, and T.D Wager. Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, page E4E7, 2023. https://doi.org/10.1038/s41586-023-05745-x.

[20] B. Tervo-Clemmens, S. Marek, R.J. Chauvin, and et al. Reply to: Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, 615:E8E12, 2023. https://doi.org/10.1038/s41586-023-05746-w.

[21] M.M. Chakravarty. Controversies on Brain-Wide Association Studies: Commentaries from the Field. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/fb60878a-5ee2-44cf-a6a9-9cdcc2f1c397.

[22] L.Q. Uddin. Brain–behavior associations depend heavily on user-defined criteria. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/5ba14033-72bb-4915-81a3-fa221302818a.

[23] P.A. Bandettini. Lost in transformation: fMRI power is diminished by unknown variability in methods and people. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/725139d7-0b8a-49dc-a81d-ba2ca64ff6d9.

[24] T. White. Behavioral phenotypes, stochastic processes, entropy, evolution, and individual variability: Toward a unified field theory for neurodevelopment and psychopathology. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/c900ce20-3ffd-4545-8c15-3ec532b2ee3b.

[25] Tiego J. and Fornito A. Putting behaviour back into brainbehaviour correlation analyses. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/2f9c5854-d10b-44ab-93fa-d485ef5b24f1.

[26] B. Thirion. On the statistics of brain/behavior associations. *Aperture Neuro*, BWAS Editorials, 2022. doi:10.52294/51f2e656-d4da-457e-851e-139131a68f14.

[27] S. Valk and M.D. Hettwer. Commentary on reproducible brain-wide association studies require thousands of individuals'. *Aperture Neuro*, BWAS Editorials: 1-2, 2022. doi:10.52294/de841a29-d684-4707-9042-5bbd3d764f84.

[28] S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. The MIT Press, 2023. http://www.fairmlbook.org.

[29] T. He, R. Kong, A.J. Holmes, M. Nguyen, M.R. Sabuncu, S.B. Eickhoff, D. Bzdok, J. Feng, and B.T.T. Yeo. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, 206:116276, 2020.

[30] T. He, L. An, P. Chen, and et al. Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. *Nat Neurosci*, 25:795–804, 2022. https://doi.org/10.1038/s41593-022-01059-9.

[31] A. Llera, T. Wolfers, P. Mulders, and C.F. Beckmann. Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *eLife*, 8:e44443, 2019. doi:10.7554/eLife.44443.

[32] U. Pervaiz, D. Vidaurre, M.W. Woolrich, and S.M. Smith. Optimising network modelling methods for fMRI. *Neuroimage*, 211, 2020. https://doi.org/10.1016/j.neuroimage.2020.116604.

[33] D.S. Bassett, P. Zurn, and J.I. Gold. On the nature and use of models in network neuroscience. *Nat Rev Neurosci*, 19:566–578, 2018. https://doi.org/10.1038/s41583-018-0038-8.

[34] S.E. Petersen and O. Sporns. Brain networks and cognitive architectures. *Neuron*, 88(1):207–219, 2015. doi:10.1016/j.neuron.2015.09.027.

[35] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, and G. Varoquaux. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 192:115–134, 2019. https://doi.org/10.1016/j.neuroimage.2019.02.062.

[36] D. Scheinost, S. Noble, C. Horien, A.S. Greene, E.M. Lake, M. Salehi, S. Gao, X. Shen, D. O'Connor, D.S. Barron, S.W. Yip, M.D. Rosenberg, and R.T. Constable. Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage*, 193:35–45, 2019. doi:10.1016/j.neuroimage.2019.02.057.

[37] G. Poldrack, R.A.and Huckins and G. Varoquaux. Establishment of best practices for evidence for prediction. *JAMA Psychiatry*, Nov 2019. doi:10.1001/jamapsychiatry.2019.3671.

[38] S. Marek, B. Tervo-Clemmens, F.J. Calabro, and et al. Towards reproducible brain-wide association studies. *bioRxiv*, 2020. https://doi.org/10.1101/2020.08.21.257758.

[39] M.P. Milham, J. Vogelstein, and T. Xu. Removing the Reliability Bottleneck in Functional Magnetic Resonance Imaging Research to Achieve Clinical Utility. *JAMA Psychiatry*, 2021. doi:10.1001/jamapsychiatry.2020.4272.

[40] R. Leenings, N.R. Winter, U. Dannlowski, and T. Hahn. Recommendations for machine learning benchmarks in neuroimaging. *Neuroimage*, 257, 2022. https://doi.org/10.1016/j.neuroimage.2022.119298.

[41] G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In Jiang T., Navab N., Pluim J.P.W., and Viergever M.A. (eds), editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI2010*, volume vol 6361 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2010. https://doi.org/10.1007/978-3-642-15705-9_25.

[42] G. Lohmann, E. Lacosse, T. Ethofer, V.J. Kumar, K. Scheffler, and J. Jost. Predicting intelligence from fMRI data of the human brain in a few minutes of scan time. *biorxiv*, 2021. doi: https://doi.org/10.1101/2021.03.18.435935.

[43] L.I. Kuncheva, J.J. Rodriguez, C.O. Plumpton, D.E.J. Linden, and S.J. Johnston. Random Subspace Ensembles for fMRI Classification. *IEEE Transactions on Medical Imaging*, 29(2):531–542, 2010.

[44] R. Bryll, R. Gutierrez-Osuna, and F. Quek. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 36(6):1291–1302, 2003.

[45] S Wold, M. Sjöström, and L Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Lab Systems*, 58(2):109–130, 2001.

[46] A. Krishnan, L.J. Williams, A.R. McIntosh, and H. Abdi. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage*, 56:455–475, 2011.

[47] S. de Jong. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Lab Systems*, 18:251–263, 1993.

[48] A.J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans on Mathematical Software*, 3:253–256, 1977.

[49] GSL – GNU Scientific Library. http://www.gnu.org/software/gsl.

[50] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform*, 8, 2014. doi:10.3389/fninf.2014.00014.

[51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[52] E. Dhamala, K.W. Jamison, A. Jaywant, S. Dennis, and A. Kuceyeski. Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. *Human Brain Mapp*, 42(10), 2021. https://doi.org/10.1002/hbm.25420.

[53] M.F. Glasser, S.M. Smith, D.S. Marcus, J.L. Andersson, E.J. Auerbach, T.E. Behrens, T.S. Coalson, M.P. Harms, M. Jenkinson, S. Moeller, E.C. Robinson, S.N. Sotiropoulos, J. Xu, E. Yacoub, K. Ugurbil, and D.C. Van Essen. The human connectome project's neuroimaging approach. *Nature Neuroscience*, 19:1175–1187, 2016.

[54] L. Snoek, M.M. van der Miesen, T. Beemsterboer, A. Van Der Leij, A. Eigenhuis, and H.S. Scholte. The amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. *Scientific Data*, 8(85), 2021. https://doi.org/10.1038/s41597-021-00870-6.

[55] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B.S. Khundrakpam, J.D. Lewis, Q. Li, M. Milham, C. Yan, and P. Bellec. The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. In *Neuroinformatics*, Stockholm, Sweden, 2013. http://preprocessed-connectomes-project.org/abide/.

[56] A. Mensch, G. Varoquaux, and B. Thirion. Compressed online dictionary learning for fast resting-state fMRI decomposition. In *Symposium on Biomedical Imaging (ISBI)*, page 12821285, 2016. doi:10.1109/ISBI.2016.7493501.

[57] R. Amthauer, B. Brocke, D. Liepmann, and A. Beauducel. Intelligenz-Struktur-Test. Technical report, Hogrefe, 2001.

[58] R.M. Lawrence, E.W. Bridgeford, P.E. Myers, and et al. Standardizing human brain parcellations. *Sci Data*, 8(78), 2021. https://doi.org/10.1038/s41597-021-00849-3.

[59] B.A. Seitzman, C. Gratton, S. Marek, R.V. Raut, N.U.F. Dosenbach, B.L. Schlaggar, S.E. Petersen, and D.J. Greene. A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *Neuroimage*, 206(116290), 2020. https://doi.org/10.1016/j.neuroimage.2019.116290.

[60] M. Joliot, G. Jobard, M. Naveau, N. Delcroix, L. Petit, L. Zago, F. Crivello, E. Mellet, B. Mazoyer, and N. Tzourio-Mazoyer. AICHA: An atlas of intrinsic connectivity of homotopic areas. *J Neurosci Methods*, 254:46–59, 2015. doi:10.1016/j.jneumeth.2015.07.013.

[61] M.F. Glasser, T.S. Coalson, E.C. Robinson, C.D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C.F. Beckmann, M. Jenkinson, S.M. Smith, and D.C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536:171–178, 2016. https://doi.org/10.1038/nature18933.

[62] A. Schaefer, R. Kong, E.M. Gordon, T.O. Laumann, X.N. Zuo, A.J. Holmes, S.B. Eickhoff, and B.T.T. Yeo. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex*, 28(9):3095–3114, 2018. doi:10.1093/cercor/bhx179.

[63] A. Di Martino, CG. Yan, Q. Li, and et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry*, 19:659–667, 2014. https://doi.org/10.1038/mp.2013.78.

[64] N.U.F. Dosenbach and et al. Prediction of Individual Brain Maturity Using fMRI. *Science*, 329(5997):1358–1361, 2010. doi:10.1126/science.