Predicting Fluid Intelligence using Areal Features derived from the Human Connectome Project Multimodal Parcellation

Varsha Vijay, Logan Z. J. Williams & Emma C. Robinson

***Abstract*—Advances in neuroimaging analyses have facilitated pivotal neuroscience comprehension, particularly when examining mechanisms underlying neurological pathologies and predicting phenotypes based on neural functional organisation. Prior studies aiming to predict fluid intelligence using neuroimaging data have been limited by variability in configuration of cortex organisation across individuals. Additionally, many of these studies focused on one modality however research demonstrates benefits to considering multiple, complementary neurobiological properties. Recently, a multimodal cortical parcellation was generated which more uniquely maps individual patterns of cortical organisation. In this paper we aim to improve performance and interpretability of machine learning models predicting fluid intelligence and overcome the mentioned limitations by harnessing this novel parcellation. The individualised cortical parcellations and data used are based on multimodal cortical surface measures acquired for the Human Connectome Project. Hand-engineered features derived from this data were deployed in nested cross-validated training of various classical machine learning models, importantly regressing out basic and non-linear confound variance. Our results demonstrate that this combination of subject-specific parcellation and multimodal neuroimaging data generated biologically meaningful features to train state-of-art predictive models of fluid intelligence. The trained stacked generalisation model outperforms deep-learning transformers and unimodal feature-based models in literature. Feature selection was additionally applied to visualise salient regions in determining fluid intelligence. Regions highlighted include the insula and superior temporal sulcus which is consistent with extant literature stating their importance in cognition. Ultimately, these results strongly encourage further work leveraging multimodal neuroimaging data and classical machine learning frameworks with a clinical cohort to detect reliable biomarkers.**

***Index Terms*—Fluid Intelligence, Human Connectome Project, Cortical Parcellation, Confound Regression**

V.Vijay, E.C. Robinson and L.Z.J. Williams are with the Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, St Thomas' Hospital, London SE1 7EH, UK. E.C. Robinson is also with the Centre for the Developing Brain, Department of Perinatal Imaging and Health, School of Biomedical Engineering and Imaging Sciences, King's College London, London, SE1 7EH, UK.

I. INTRODUCTION

THE increasing prevalence of neurological and neurodevelopmental disorders has been accompanied by diverse efforts to map brain organisation in health and disease [1]. This is partially attributed to functional brain organisation measures gaining recognition as effective clinical biomarkers in neuropsychiatric disorders [2]. For instance, functional connectivity (FC) has been studied in numerous investigations which have demonstrated that FC measures are altered in autism spectrum disorder (ASD) [3], depression [4] and Alzheimer's [5]. Moreover, functional organisation metrics have been successfully adopted to examine variability in phenotypes [1, 6, 7].

Cognitive ability is one such phenotype, however the exact mechanisms determining it are still poorly understood. It is biomedically significant to explore the functional configurations associated with cognition; these can be harnessed to provide indicators of cognitive decline which is characteristic of Alzheimer's and other neuropsychiatric diseases [8, 9]. As such, an accurate pipeline exploiting inter-subject deviations in functional brain organisation to predict cognitive ability would prove highly advantageous; appropriate prevention or timely interventional treatments could be adopted. There have been varied attempts in neuroimaging literature to achieve this goal however, the majority of these have not capitalised on the advantages of considering multiple neuroimaging modalities. Furthermore, many prior studies delineated regions harnessing group average maps which fail to account for individual variability.

The predominant aims of this paper were to overcome these limitations by using subject-specific parcellations and multimodal cortical data to create a more reliable and interpretable machine learning model to predict fluid intelligence, and to identify salient cortical regions in the context of cognition.

A. Design Statement

In this paper, a novel approach is taken to achieve the aim of hand-engineering biologically meaningful, areal-based features to train a variety of classical machine learning models and establish their potential in predicting fluid intelligence. Specifically, the features were created by combining multimodal cortical surface data, made available by the Human Connectome Project (HCP) [10], with individual subject HCP multimodal parcellations [11] which more uniquely map

individual patterns of cortical organisation. A significant distinction between this study, and previous approaches to elicit underlying macroscopic brain systems, is that multiple cortical surface measures are considered here in predicting fluid intelligence. Furthermore, the variance due to confound variables were accounted for via cross-validated confound regression, ensuring any demonstrated fluid intelligence variance was more likely to be due to differences in functional organisation of the brain.

The main contributions presented in this paper include:

- An optimised machine learning algorithm to predict fluid intelligence which outperforms previous attempts.
- Identification of cortical regions important for predicting fluid intelligence and cognitive function.
- A pipeline accounting for variation due to confound variables including age, sex, weight and intracranial volume.
- An approximation of the contribution of resting-state functional MRI (rfMRI) data to the overall multimodal data in its ability to predict fluid intelligence.

The timescale for this study was ~3 months and was used to produce deliverables in 4 phases:

1. A literature review was first conducted to understand the strengths and limitations of previous approaches.
2. Two areal feature sets were engineered: one without deconfounding and one with deconfounding to systemically investigate the effects of applying cross-validated confound regression.
3. Various classical machine learning algorithms were trained on the feature sets.
4. rfMRI data for the subjects were also pre-processed into correlation matrices, producing features to train machine learning models.

A detailed timeline with respect to deliverables can be found in Appendix I.

B. Background

Typical cognitive functioning is reliant on complex operations and coordination across distinct functional brain systems [12, 13].

1) Limitations of Univariate Analyses

Recently, there has been a research emphasis on investigating statistical associations between cognitive phenotypes and these brain systems. Variability in cognition has mainly been investigated for its correlation with particular brain measures including structural connectivity [14, 15], task-based activity [16], cortical thickness [17] and resting-state FC [18]. The analyses of single neuroimaging variables, however, do present drawbacks. Firstly, such studies do not consider mutual inter-feature associations in the brain, overlooking redundant variance sources. Secondly, statistical power is reduced due to accounting for multiple testing and false positives are more likely to occur in mass univariate studies [1]. Furthermore, by utilising a univariate approach to examining cognitive ability, this decreases potential clinical applications and generalisability of trained predictive models.

Many studies have benefitted from using multivariate data to train machine learning algorithms to overcome the mentioned limitations. Literature harnessing this technique to predict cognitive ability has focused heavily on rfMRI because of its multivariate nature [19]. In comparison to univariate analyses, these machine learning based models reveal fundamental insights into biomarkers of cognitive ability, such as the algorithm's generalisability [1], motivating the selection of this approach in this study to increase likelihood of neurobiologically meaningful findings.

2) HCP Multimodal Functional Organisation Data

Prior investigations into the relationship between fluid intelligence and cortical organisation have focused on a single neuroimaging modality. Literature has evidenced the advantages in studying multiple modalities; diffusion and functional MRI (fMRI) aid in demonstrating distinct properties of underlying brain systems which are separately linked with various components of cognition. By utilising a multimodal approach, as recent studies have done [1, 11], more neurobiological characteristics can be assessed for their predictive value. In the current study, cortex topography, connectivity, function and architecture are neurobiological properties examined by using multimodal data from the HCP. The task-based fMRI provided insight into cortical function [20], rfMRI highlighted FC across and the topography within cortical regions [11] and T1 and T2 weighted structural MRI delineated architectural metrics including cortical thickness and myelin content [21, 22]. One critical limitation resulting from analysing multimodal cortical data is that in consolidating the multiple data sources, data dimensionality and model complexity are dramatically increased which can lead to model overfitting [23]. One method used to overcome this issue was to hand-engineer features by combining the multimodal data with a novel cortical parcellation.

3) HCP Multimodal Cortical Parcellation Map

A cortical parcellation is essential in neuroimaging to facilitate communication of findings, to enable more neurobiologically meaningful results and to reduce data dimensionality, increasing statistical power [11]. Previously developed cortical maps have distinguished between 50-200 parcels [24, 25] based on unimodal data. Glasser et al. uniquely generated an areal parcellation based on multimodal data from the HCP, segmenting the entire cortex into 360 regions [11].

A semi-automated technique was used to determine gradients in the neurobiological properties: possible regional boundaries were identified algorithmically which were then studied by neuroanatomists. A machine learning classifier was then trained to locate each region in unseen individual subjects employing distinct areal 'fingerprints' [11]. The novel contribution of this paper was the final parcellation's ability to determine regions successfully in subjects exhibiting variation in parcels. Therefore, the utility of the HCP multimodal cortical mapping in this study was motivated by its novel sensitivity to atypical

parcellations and the more meaningful results obtained via neurobiologically defined regions.

4) Biomedical Significance of Fluid Intelligence

Understanding factors influencing cognition and behaviour has been an extant goal to bridge research from psychological and biological standpoints [26]. Fluid intelligence is one such cognitive phenotype that demonstrates inter-subject variance and has been studied in many neuroimaging investigations. The concept of fluid intelligence refers to problem-solving and abstract reasoning abilities [27], in contrast to crystallised intelligence which depicts accumulated knowledge through an individual's experiences. The HCP dataset was also composed of phenotypic metrics including Penn Matrix Test (PMAT) scores and these were used as quantitative inferences of fluid intelligence.

There is substantial biomedical interest to comprehend the biological mechanisms that give rise to variations in fluid intelligence; many neurological disorders are correlated with rapid declines in fluid intelligence [28]. Furthermore, to enable effective interventional treatments, it is increasingly important to identify at-risk individuals and diagnose neurological disorders early on so preventive and interventional measures can be established. Therefore, assessing inter-subject heterogeneity in neurobiology, potentially causing differences in fluid intelligence, would be beneficial in detecting clinical biomarkers.

5) Socioeconomic Context

There is a growing population globally with neurological disorders and systematic reviews have found stroke and Alzheimer's disease place the biggest strain on the US healthcare system [28]. It is approximated globally that 1 in 14 adults over 65 have been diagnosed with dementia [29]. Moreover, neuropsychiatric disorders can co-occur [30], causing compounded socioeconomic impacts for patients.

The quality of life for individuals suffering from developmental disabilities and their families can be negatively impacted due to employment discrimination [31, 32] and decreased levels of social support, often associated with worse treatment outcomes. Research into such disorders is paramount as the findings can influence education policies impacting young individuals suffering from developmental disorders [33].

Currently, diagnoses is heavily dependent on clinical observation. This causes an economic burden on healthcare systems as well as critically delaying necessary intervention for patients. The NHS 2021 report exemplifies this, stating that the average waiting time before diagnoses of ASD was 292 days [34]. In the US, the economic burden of neurologic disease is estimated at $800 billion per year and approximately €134 billion annually in the UK [35]. Consequently, it is paramount to investigate effective clinical biomarkers that can identify at-risk individuals as well as facilitate automation of the current diagnostic process. Commercially, this would be widely beneficial in saving resources and improving the patient pathway for neurodegenerative conditions.

6) Potential Ethical and Legal Considerations

There are paramount legal and ethical considerations when machine learning methods are introduced in computational neuropsychiatry. These include data protection and fairness as well as transformative impacts of machine learning applications.

Potential biomarkers highlighted in this paper's results would need to be verified in a clinical cohort. Accurate clinical indicators could give rise to innovative treatments that depend on reliable biomarkers. For instance, results could aid the development of deep brain stimulation intervention in epilepsy [36], a prevalent neurological disorder. Thorough clinical trials would be required to determine any safety risks however trials present ethical issues due to the risks faced by participants. Any resulting treatments may also pose legal accountability issues for consequences in the event of treatment failures.

7) Accounting for Confound Variables

A fundamental limitation in assessing the association between cortical properties and phenotypic heterogeneity is ambiguity as to which information source is truly driving variance in the target. In this paper, there are confounds which exhibit variance that could be linked to differences in fluid intelligence such as age. For instance, Nooyens et al. found differences in cognitive functioning with aging between sexes [37]. Proposed techniques to isolate the variance in a psychometric target caused by specific features include confound regression [38] and post hoc counterbalancing [39]. Confound regression is more frequently deployed and functions by subtracting confound variance directly from the features. A study into the effectiveness of these methods was conducted by Snoek et al. where the findings indicated that both methods caused heavy bias in results [40]. Importantly, this study also concluded that the negative bias exhibited with confound regression could be ameliorated using a cross-validated form, deeming it the most appropriate method to account for confound variance [40]. Therefore, cross-validated confound regression was adopted in the current study to more effectively elucidate variance in cortical surface properties influencing fluid intelligence heterogeneity.

II. METHODS

The computational methods described were implemented in Python3 leveraging libraries Sci-kit learn [41], Nibabel and Pandas. The entire project code can be found at the GitHub link: https://github.com/varsha-vijay/HCP-IQ-Project.

A. Data and Acquisition

The data utilised in this project are multimodal cortical data from 446 subjects from Glasser et al. [11] and were acquired using high resolution 3T MR scans. The multimodal data consisted of 112 distinct metric files describing various functional MRI and cortical structure metrics including cortical folding, thickness, myelin content and resting and task-based fMRI. The dataset summarised these measures for approximately 30,000 different vertices across the cortical

surface for each brain hemisphere, using the FS_LR32K space [10]. The rfMRI timeseries data for each subject comprised of 4 recordings acquired over 1200 timepoints with intervals of 0.72 seconds, providing a total recording time of ~1 hour.

Importantly, the HCP data release featured behavioural measures including PMAT fluid intelligence scores which were utilised as target variables for the training and evaluation of machine learning models.

B. Data Pre-processing and Feature Engineering

To engineer biologically meaningful feature sets of reasonable dimensionalities, the data was pre-processed and combined with the novel multimodal cortical parcellation map. As mentioned, a distinct novelty of this paper is the utility of subject-specific cortical parcellations which delineate 360 regions, with high sensitivity to individual variation of features.

1) Multi-Modal Cortical Surface Data

Features were created for the 112 cortical metric files by extracting all measures within a parcellation and taking the regional average. This step yields 40,320 (112x360) features for each subject. The utility of the parcellation map critically reduced the dimensionality of the data from over 60,000 vertices to just 360 regional averages. An instance of the averaged measures for one cortical metric file is visualised in fig.1. This was one of the feature sets trained on with the supervised machine learning models discussed below, without

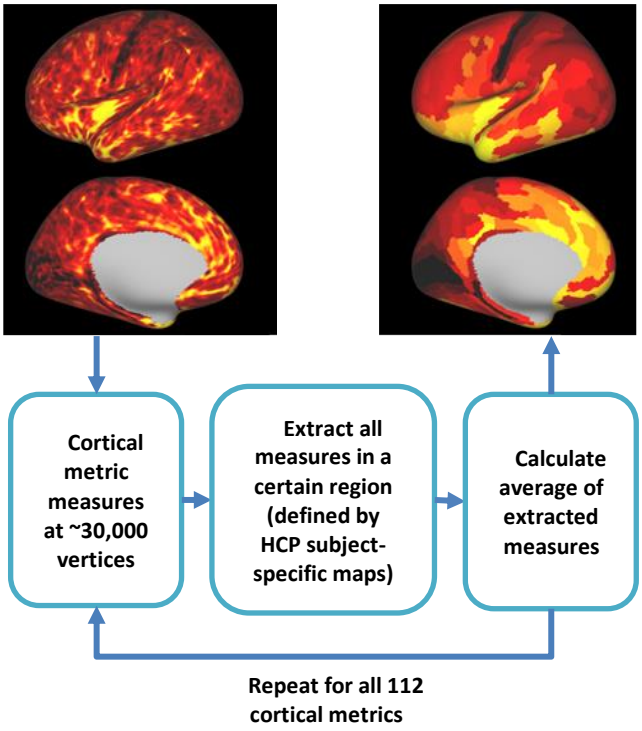

Fig. 1. Example of a cortical metric file, cortical thickness corrected for folding, represented on a group average very inflated left hemisphere. The top and bottom images of the cortex are lateral and medial views respectively. The left image represents the thickness measures for over 30,000 vertices being reduced in dimensionality by taking regional averages defined by the multi-modal cortical map parcellation, seen on the right images. This process is repeated for all 112 cortical metrics resulting in an average map for each metric for each subject.

accounting for confounding variables, to elicit any consequences of deconfounding features on predictive performance of models.

2) Cross-Validated Confound Regression

The PMAT scores were used as the target variable to be decoded from the multimodal neuroimaging data. Cross-validated confound regression was employed to remove variance in the target caused by confound variance, following the approach harnessed by Pervaiz et al. [2], originally proposed by Snoek et al. [40]. Certain variables were identified to be confounds including age, weight, sex and intracranial volume. Aside from these basic confounds, some non-linear impacts were additionally considered including age^2 and sex x age^2 to further reduce ambiguity of the relative contributions of both the confound and true signal [2]. To obtain a plausible performance and better comprehension of the underlying neuroimaging data sources causing variation in fluid intelligence, the cross-validated approach was used [40].

For each fold in the cross-validation model, a linear regression model was fitted onto the feature set, deploying the confound variable as the predictor [40]. Equation (1) below depicts the relationship modelled with the confounding variables matrix C, an approximated parameter β, each feature j in the feature matrix X and error ε.

$$Xj = C\beta + \varepsilon \tag{1}$$

Least squares was then leveraged to estimate the parameter β as listed in the matrix calculation (2).

$$\beta j = (C^T C)^{-1} C^T \, Xj \tag{2}$$

The final deconfounding step involved regressing out the confound variance from the data features by subtracting confound data from the primary feature set [40] as seen in (3), where Xj,corr refers to the final corrected feature matrix.

$$Xj,corr = Xj - C\beta j \tag{3}$$

Ultimately, the yielded matrix Xj,corr will have the confound variance and its variance with the target variable subtracted, hence the contributions of the true signal will theoretically be more interpretable from the predictive models. This allows a more critical examination of the true association between cortical organisation and heterogeneity exhibited in fluid intelligence. This computation was implemented for the training, validation and test features for each fold.

3) Resting-State Functional MRI Timeseries Data

The rfMRI data contained BOLD signals across considered brain vertices for every timepoint. The original file formats were CIFTI types, containing data for both hemispheres so firstly, a bash script using the wb_command function *CIFTI Separate* was deployed, dividing each of the four files into corresponding right and left hemisphere readings. Iterating over each subject, the measurements for all vertices in a region, defined by the subject-specific parcellations, were averaged. This obtained 360 measurements for each of the 1200 timepoints for four different recordings. These were then concatenated to generate a single timeseries per region and used to generate a 360x360 Pearson's correlation matrix for each subject, an example of which is seen in fig.2. This correlation matrix showed insight into the static FC for each subject. As the derived correlation matrices are symmetric, the matrix diagonal and lower triangle contain redundant information and therefore were discarded. Each upper triangular value was stored to form the feature vectors for every subject. Consequently, each subject had 64,620 features.

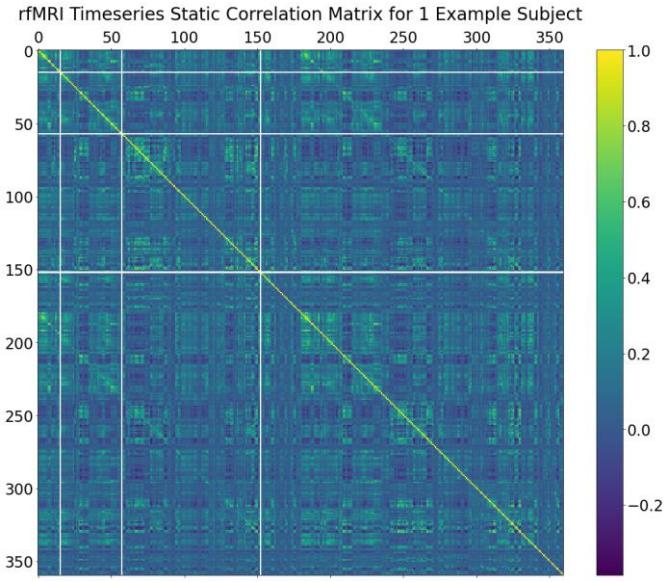

Fig. 2. An exemplar static correlation matrix for one subject's rfMRI timeseries. The matrix is of dimensions 360 x 360, corresponding to each of the 360 delineated regions across both hemispheres by the HCP individual cortical parcellation maps. Only the upper triangle was extracted to provide features for each subject due to the matrix symmetry. The white lines correspond to invalid recording measures for certain areas which were omitted from the final feature matrices.

C. Nested Cross-Validation Folds and Selection

During the training of the machine learning models, an integrated loop was used during cross-validation so that the nested extension could be applied. This approach was motivated by an attempt to avoid overfitting by optimising the hyperparameters discussed later.

To increase the compatibility with and performance of the trialled Scikit-learn machine learning models, the features were scaled and centered around a mean of 0 and a variance of 1. The subjects were split into 5 distinct folds to ensure there was no family leakage across the folds, reducing the chances of an inaccurate sense of model generalisation. For instance, the subject group contained both identical and fraternal twins and if a model was trained on a twin and then tested on their sibling, the validity of the results would be compromised. The five splits were organised into training, test and validation groups as shown in table 1.

Folds	Training	Test	Validation
1	S3,S4,S5	S1	S2
2	S1,S4,S5	S2	S3
3	S1,S2,S5	S3	S4
4	S1,S2,S3	S4	S5
5	S2,S3,S4	S5	S1

Table. 1. Summary of how each split, S, was organised to provide the training, test, and validation groups for each fold.

Accordingly, for each fold, the models were adjusted on the training dataset and then evaluated on the unseen test set. This was achieved by computing the error between the predicted and known labels of the test sets. Various error measurements can be leveraged to compare models and, in this case, the R^2 score was selected. This metric choice was justified as it is an appropriate measure in regression to quantify the variance in the dependent variable that can be attributed to the variance in the independent variables. Furthermore, the correlation between model predictions and the ground truth labels could be simply calculated by taking the square root of the R^2 score. Previous attempts predicting cognitive ability have employed correlation to evaluate model performance [1, 31]. Therefore, the computed R^2 scores provided an efficient way to determine correlations for comparison with literature. Using the nested cross-validated approach, the R^2 scores were averaged across all 5 folds and recorded.

The hand-engineered feature matrices still possessed large dimensionality, increasing the likelihood of overfitted models after training; feature selection was mostly performed before training using in-built Random Forest feature importances. Feature selection was an advantageous step as it also decreased the computational burden, required system memory and running time.

D. Optimising Machine Learning Models

A range of different supervised regression models were trained on the engineered features. These included penalised regression models, such as Lasso regressors, and ensemble-based methods like Random Forest and AdaBoost regressors. Importantly, the final model trained and evaluated was a meta-learning stacking framework. A schematic of this framework can be seen in fig.3. These algorithms built predictions based on multiple base machine learning models. Different selections of base models were evaluated and the optimum combination was used for each feature set. An exhaustive list of the different machine learning frameworks trained is detailed in table 2 along with the respective hyperparameters optimised for. The error metric chosen in this project when training models was the mean absolute error as it is less impacted by outliers compared to measures such as the root mean squared error.

E. Visualising Salient Neurobiological Regions

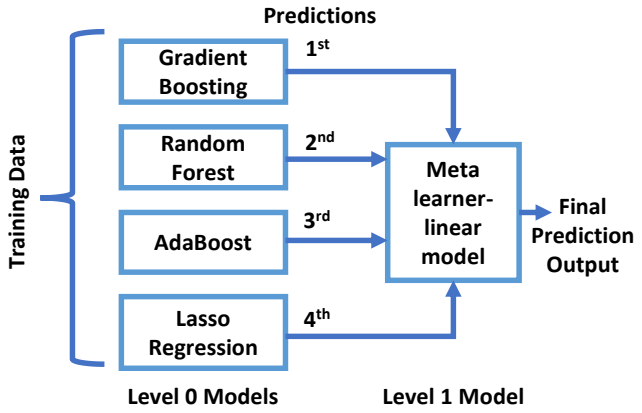

Fig. 3. A schematic diagram representing the two different levels in the meta-learning stacked generalisation model. This specific example was applied to the multimodal cortical surface features before deconfounding. Each of the base level 0 models (Gradient Boosting, Random Forest, AdaBoost and Lasso Regression) are trained first on the training data, finding the optimum hyperparameters. Each of these optimal models then feed their predictions to the meta learner level 1 linear model which outputs the final prediction.

Machine Learning Model	Hyperparameters Optimised
Gradient Boosting	Minimum samples per leaf, max features determining best split
Random Forest	Minimum samples per leaf, max features determining best split
AdaBoost	Loss function, maximum number of estimators at which boosting stops
Lasso Regression	Alpha: regularisation hyperparameter
Stacking	

Table. 2. Listing the models trained as well as the hyperparameters optimized for each model as part of the nested cross-validation.

The most important neural regions in determining predictive intelligence were additionally visualised for the multimodal feature set and the deconfounded feature set for comparison. A nested-cross validation training approach was used for a Random Forest regressor. The importances for each feature were extracted and averaged across the folds. Subsequently, the features were correlated and traced back to the multimodal parcellation region they belonged to. The importances were then averaged for each parcel providing 360 different regional importances (180 per hemisphere). The 100 highest importances determined the 100 most salient regions in influencing variation in fluid intelligence. These regions were then visualised in FS_LR32K space for the multimodal data and deconfounded multimodal data using Connectome Workbench.

III. RESULTS

A. Multi-Modal Cortical Surface Data

1) Machine Learning Model Performances

The R^2 scores for models trained on the multimodal cortical surface features before deconfounding are presented in table 3. The models were trained on different numbers of the initial features including the total, a half, and a quarter of the original features. To identify which features were eliminated before training to result in a sub-selection of features, the Random Forest regressor importances were treated as deciding factors. The results demonstrate that the stacked model performed best consistently compared to the other models, followed by Lasso regression. Before deconfounding, the stacked model was trained using the other four optimised models as base models for meta-learning. The below models were previously trained using the same data before deconfounding in a preliminary investigation, by the same research group, and the same R^2 scores are achieved here, demonstrating the repeatability of the results.

Feature Number	**40,320 (All)**	**20,160 (1/2)**	**10,080 (1/4)**
Trained Model	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**
Gradient Boosting	0.0869 ± 0.007	0.0770 ± 0.004	0.0687 ± 0.008
Random Forest	0.0410 ± 0.001	0.0649 ± 0.0003	0.0877 ± 0.002
AdaBoost	0.0139 ± 0.02	0.0720 ± 0.02	-0.00906 ± 0.02
Lasso	0.109 ± 0.06	0.120 ± 0.01	0.141 ± 0.07
Stacked	**0.153 ± 0.007**	**0.124 ± 0.01**	**0.153 ± 0.005**

Table. 3. The mean R^2 scores achieved across all folds for the multimodal cortical surface features before deconfounding. The R^2 scores for the best performing model, stacked generalisation, are shown in bold.

After applying cross-validated confound regression to the multimodal cortical surface features, the R^2 scores obtained were lower but comparable for each trialled model. These are captured in table 4. AdaBoost regression did not perform well for the deconfounded features, consistently returning negative R^2 scores. Consequently, it is not included in the results table and was also excluded as a base model from the stacked generalisation algorithm. After accounting for the effect of confound variables, the stacked model still performs best for all and half the number of features selected. The best performance is achieved after feature selecting for half the original features with the stacking model, as highlighted in table 4. The variances are consistently low across both models and number of features selected in training, demonstrating stable performances of the models across folds, both before and after deconfounding.

Feature Number	40,320 (All)	20,160 (1/2)	10,080 (1/4)
Trained Model	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**
Gradient Boosting	0.0257 ± 0.005	0.0611 ± 0.004	0.0416 ± 0.0008
Random Forest	0.0543 ± 0.0006	0.0460 ± 0.0009	0.0672 ± 0.001
Lasso	0.0621 ± 0.006	0.0405 ± 0.01	0.0621 ± 0.0002
Stacked	**0.0666 ± 0.002**	**0.0918 ± 0.004**	**0.0624 ± 0.008**

Table. 4. The mean R^2 scores achieved across all folds for the multimodal cortical surface features after applying cross-validated confound-regression. The R^2 scores for the best performing model, stacked generalisation, are shown in bold.

2) Visualising Salient Cortical Regions

The 100 most significant regions are visualised in fig.4A and fig.4B respectively. Notably, there appears to be a level of symmetry in the salient regions highlighted both before and after accounting for confounds. This indicates that the same areas from both hemispheres appear to be contributing to the prediction of fluid intelligence scores. Specifically, the insula and superior temporal sulcus are highlighted in both figures. Before deconfounding, fig.4A also demonstrates the identification of somatosensory and visual regions as salient regions however these are removed after deconfounding and are not visualised in fig.4B.

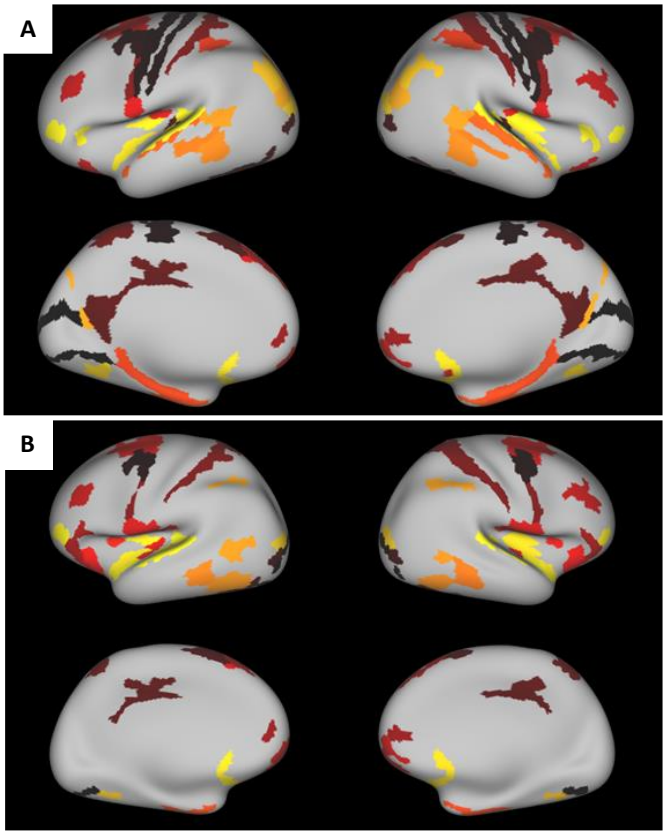

Fig. 4A & B. Visualising the 100 most significant regions on the left and right hemispheres in predicting fluid intelligence before deconfounding (A, top) and after applying cross-validated confound regression on features for the multimodal cortical surface data (B, bottom). Features are visualised on group average very inflated surfaces.

B. Resting-State Functional MRI Timeseries Data

The models trained on the features collated using the rfMRI data were used as examples of models created using unimodal data in comparison to the other feature sets generated in this study and the best performing model was the Random Forest regressor. Due to a very large number of features (over 60,000), the feature selection provided half and quarter the number of features to train on. The R^2 scores from these optimised models are shown in table 5. As seen in this table, the R^2 score achieved for the model trained on half the number of features was noticeably higher.

Feature Number	32,310 (1/2 of Feature Set)	16,155 (1/4 of Feature Set)
Trained Model	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**	**Mean R^2 ± Variance (3s.f. ± 1s.f.)**
Random Forest	0.0390 ± 0.001	0.0272 ± 0.001

Table. 5. The mean R^2 scores achieved across all folds for the Random Forest model trained on rfMRI timeseries correlation features.

C. Evaluation of Machine Learning Model Performance

The correlations of the best performing models in this study are listed with correlations achieved using different fluid intelligence prediction algorithms [1, 31] in table 6. Before deconfounding the multimodal cortical surface data, the best performing model was the stacked algorithm which outperformed previous attempts to predict fluid intelligence in literature, leveraging transformer models [31] and alternate stacking approaches [1]. After cross-validated confound regression, the correlation score for stacking was marginally lower. All correlation scores achieved in this study outperform the stacking model trained on multimodal data by Rasero et al. [1] and are comparable to literature benchmarks.

	Correlation (2.s.f)
Stacked model: before deconfounding data	0.39
Stacked model: after deconfounding data	0.30
Random forest regressor: RS-fMRI data	0.20
Transformer Model (Dahan et al.)	0.35
Stacking (Rasero et al.)	0.19

Table. 6. The correlations from the current study's best performing models compared with correlations achieved using different prediction algorithms in recent research.

IV. DISCUSSION

The results demonstrate that the stacking models trained on neurobiologically meaningful features are state-of-art with the highest correlation scores in recent literature for predicting fluid intelligence. However, certain limitations must also be considered when interpreting the findings here.

A. Interpretation of Results

1) Machine Learning Model Performances

Table 3 demonstrates the consistently high performance of stacking. This result supports ensemble theory which states that base model heterogeneity in a meta-learning model is advantageous; they are more likely to identify unique information in training data. Interestingly, the R^2 score decreases when using half the features selected but the performance improves after a further round of feature selection. The discarding of noisy features could explain this increased R^2 score after further feature selection.

With the deconfounded feature set, table 4 shows that the other models seem to have more comparable performances with the stacking model, confirming that nested cross-validated confound regression is an efficient and appropriate technique to account for confound variance. Overall, the best stacking model has a slightly lower performance after deconfounding which may be attributed to the excluded variance unrelated to the cortical surface data. Regardless, the deconfounding step is massively beneficial as it helps highlight the true association between functional organisation and fluid intelligence. Further interpretation of these results can be found in Appendix II.

The optimised stacking models have very similar performances to the deep learning surface vision transformer model generated by Dahan et al. [31] as evidenced in table 6. Complex deep learning frameworks become increasingly difficult to interpret [42] and require high volumes of data and computational power to ensure generalisability, however these are not always available. Therefore, these results elicit how classical machine learning models may present a useful option when deep learning approaches are not feasible due to lack of substantially sized biomedical datasets and computational requirements.

The stacking model proposed by Rasero et al. [1] is a further valid study for comparison as they also used multimodal imaging data to predict fluid intelligence. The key distinctions between these studies is that Rasero et al. had a greater dataset of 1028 subjects and employed group average parcellations instead of subject-specific mappings. Comparison of the papers' findings clearly reflects the advantages of using a cortical map more sensitive to individual cortical variation as the derived stacking models here outperform theirs, even with a smaller dataset.

The Random Forest model trained on rfMRI data marginally outperformed the stacking model with a much smaller dataset, further highlighting how advantageous the HCP subject-specific maps were to the performance of the models. Additionally, when only considering one modality, rfMRI, the performance of the models is decreased, providing evidence corroborating the benefits to considering multiple complementary neurobiological properties in training predictive models.

2) Visualisation of Salient Cortical Regions

The visualisation of the most important regions in predicting fluid intelligence highlighted certain areas associated with cognition in prior literature. For instance, there have been many studies discussing the fundamental roles of the insula in cognition [43, 44]. Both fig.4A and fig.4B identify regions around the insula as salient in determining fluid intelligence which is in accordance with its suggested role in cognitive function. Before deconfounding, somatosensory and visual regions are additionally captured in the visualisation in fig.4A however these are not highlighted in fig.4B. This finding is expected, as these regions are less relevant to cognitive science, and reflects the appropriateness of cross-validated confound regression.

Importantly, there is persistence of regions surrounding the superior temporal sulcus which is widely perceived to subserve for cognition [45-47]. Therefore, this result also aligns with extant literature. A further notable point about the visualisations is that both hemispheres before and after deconfounding are roughly symmetric, yet unidentical. These findings suggest that similar regions from both hemispheres perform an active role in cognition.

B. Limitations and Future Improvements

1) Training Datasets

As multimodal data was employed, the number of features was vast, making models more prone to overfitting. Therefore, the number of participants posed a limitation in this study as it was significantly less than the number of features. Future studies could leverage larger cohorts such as UK Biobank datasets [48] to enhance model generalisability. Moreover, this study's participants were primarily healthy young adults and to identify true clinical indicators of neuropsychiatric disorders, clinical and substantially representative cohorts are required.

The socioeconomic impacts of identifying reliable biomarkers for neurological disorders is massive; at-risk individuals could be identified, allowing preventive measures to be established, alleviating the socioeconomic burden on healthcare systems. Additionally, reliable indicators could help introduce more automation in diagnoses however safety and ethical considerations must be taken to ensure of informed consent and that ethical clearing is acquired. Furthermore, there is a legal question of accountability that presents itself when a prediction of fluid intelligence is made.

The lack of diversity in biomedical datasets is a further cause for ethical, legal and safety concerns. Most available neuroimaging data are from white participants, resulting in a critical bias present in almost all artificial intelligence literature, causing ethnic minorities to be discriminated against. In the context of predicting fluid intelligence, if a generated model does not translate to data from other ethnicities and the model performs poorly for certain populations, this is likely to have a detrimental impact on their health and social care, adding to discrimination they may already face.

2) Static vs. Dynamic Functional Connectivity

An additional limitation is that only static FC was considered using the rfMRI timeseries data, meaning temporal FC variation was not considered. Many studies have demonstrated how the brain cycles through a repertoire of FC states [49], evidencing the need to study temporal FC properties. An improvement on this study would be to adopt a dynamic FC analysis pipeline such as Leading Eigenvector Dynamic Analysis [12] in which connectivity correlation matrices are yielded for every timepoint. Using these matrices to generate features in future work may result in improved predictive value of rfMRI in the context of fluid intelligence.

3) Stacked Generalisation Models

A further area worth investigating is the effects on stacking model performance when base models are each trained on a different subset of the original features. This would be advantageous as the number of features would be reduced. The novel approaches adopted here to use multimodal data with subject-specific parcellations to train classical machine learning algorithms have provided compelling evidence of their utility in predicting fluid intelligence. Future work could seek to replicate this approach for different behaviours to establish further clinical biomarkers for other neurological disorders. Identifying heterogeneity in phenotypes can further research and help classify further sub-divisions of such disorders, yielding more subject-specific diagnoses [31]. This is a clear goal in healthcare as there are already significant efforts to generate more objective criteria for diagnosing psychiatric disorders such as the Research Domain Criteria Initiative [50].

4) Identifying Salient Cortical Features

There is value in studying salient regions in future investigations, particularly considering asymmetries across hemispheres as certain cognitive functions, such as spatial processing, have been reported to be asymmetrically distributed between hemispheres [51, 52]. Consequently, further studies delving into significant regions in the context of cognition could highlight fundamental hubs contributing to specific functions. An additional extension of this work could seek to trace important features to the particular neurobiological property they are associated with and ascertain whether certain modalities hold more weighting in predicting fluid intelligence.

V. Conclusion

The findings here demonstrate that employing subject-specific parcellations with multimodal neuroimaging data improves the performance and interpretability of fluid intelligence prediction models, strongly favouring their utility in future studies predicting phenotypes. In particular, the state-of-art stacking models derived outperform previous deep learning approaches, suggesting that classical machine learning methods should not be overlooked in neuroimaging studies, especially as large, representative datasets are still lacking. Further evaluations and clinical trials could result in the detection of reliable biomarkers of cognitive decline, streamlining prevention and interventional treatments of neurological diseases.

Acknowledgment

I would like to acknowledge Dr Emma Robinson and Dr Logan Williams for their continued guidance and support throughout this project. They equipped me with the tools, knowledge and fascinating research proposal which presented many rewarding learning opportunities. In addition, I am very appreciative of the time, help and suggestions provided by Dr Mohamed Suliman, Simon Dahan, Abdul Wahab Majeed and Abdulah Fawaz during this study. The data used was provided by the Human Connectome Project, WU-Minn Consortium (David Van Essen and Kamil Ugurbil are Principal Investigators: IU54MH091657) funded by the 16 NIH Institutes and Centers supporting the NIH Blueprint for Neuroscience Research and the McDonnell Center for Systems Neuroscience at Washington University [10]. The multimodal cortical parcellation used was created by Glasser et al. [11].

References

1. Rasero, J., et al., *Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability.* PLoS Comput Biol, 2021. **17**(3): p. e1008347.
2. Pervaiz, U., et al., *Optimising network modelling methods for fMRI.* Neuroimage, 2020. **211**: p. 116604.
3. Nair, A., et al., *Impaired thalamocortical connectivity in autism spectrum disorder: a study of functional and anatomical connectivity.* Brain, 2013. **136**(Pt 6): p. 1942-55.
4. Anand, A., et al., *Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study.* Biol Psychiatry, 2005. **57**(10): p. 1079-88.
5. Greicius, M.D., et al., *Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4637-42.
6. He, B.J., et al., *Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect.* Neuron, 2007. **53**(6): p. 905-18.
7. Kelly, A.M., et al., *Competition between functional brain networks mediates behavioral variability.* Neuroimage, 2008. **39**(1): p. 527-37.
8. Hallowell, E.S., et al., *The complementary utility of cognitive testing and the medication management ability assessment in older adults.* Neuropsychology, 2022.
9. Huang, M.F., et al., *Neuropsychiatric symptoms and mortality among patients with mild cognitive impairment and dementia due to Alzheimer's disease.* J Formos Med Assoc, 2021.
10. Van Essen, D.C., et al., *The WU-Minn Human Connectome Project: an overview.* Neuroimage, 2013. **80**: p. 62-79.
11. Glasser, M.F., et al., *A multi-modal parcellation of human cerebral cortex.* Nature, 2016. **536**(7615): p. 171-178.
12. Alonso Martinez, S., et al., *The Dynamics of Functional Brain Networks Associated With Depressive Symptoms in a Nonclinical Sample.* Front Neural Circuits, 2020. **14**: p. 570583.

13. Petersen, S.E. and O. Sporns, *Brain Networks and Cognitive Architectures.* Neuron, 2015. **88**(1): p. 207-19.
14. Li, Y., et al., *Brain anatomical network and intelligence.* PLoS Comput Biol, 2009. **5**(5): p. e1000395.
15. Zimmerman, M.E., et al., *The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan.* Am J Geriatr Psychiatry, 2006. **14**(10): p. 823-33.
16. Wager, T.D., et al., *Common and unique components of response inhibition revealed by fMRI.* Neuroimage, 2005. **27**(2): p. 323-40.
17. Schnack, H.G., et al., *Changes in thickness and surface area of the human cortex and their relationship with intelligence.* Cereb Cortex, 2015. **25**(6): p. 1608-17.
18. Shen, X., et al., *Resting-State Connectivity and Its Association With Cognitive Performance, Educational Attainment, and Household Income in the UK Biobank.* Biol Psychiatry Cogn Neurosci Neuroimaging, 2018. **3**(10): p. 878-886.
19. Sui, J., et al., *Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises.* Biol Psychiatry, 2020. **88**(11): p. 818-828.
20. Barch, D.M., et al., *Function in the human connectome: task-fMRI and individual differences in behavior.* Neuroimage, 2013. **80**: p. 169-89.
21. Glasser, M.F., et al., *The minimal preprocessing pipelines for the Human Connectome Project.* Neuroimage, 2013. **80**: p. 105-24.
22. Glasser, M.F., et al., *Trends and properties of human cerebral cortex: correlations with cortical myelin content.* Neuroimage, 2014. **93 Pt 2**: p. 165-75.
23. De Martino, F., et al., *Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns.* Neuroimage, 2008. **43**(1): p. 44-58.
24. Felleman, D.J. and D.C. Van Essen, *Distributed hierarchical processing in the primate cerebral cortex.* Cereb Cortex, 1991. **1**(1): p. 1-47.
25. Nieuwenhuys, R., *The myeloarchitectonic studies on the human cerebral cortex of the Vogt-Vogt school, and their significance for the interpretation of functional neuroimaging data.* Brain Struct Funct, 2013. **218**(2): p. 303-52.
26. Kosslyn, S.M., et al., *Bridging psychology and biology. The analysis of individuals in groups.* Am Psychol, 2002. **57**(5): p. 341-51.
27. Gray, J.R., C.F. Chabris, and T.S. Braver, *Neural mechanisms of general fluid intelligence.* Nat Neurosci, 2003. **6**(3): p. 316-22.
28. Collaborators, G.U.N.D., et al., *Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study.* JAMA Neurol, 2021. **78**(2): p. 165-176.
29. *Alzheimer's Society's View On Demography*. [cited 2022; Available from: https://www.alzheimers.org.uk/about-us/policy-and-influencing/what-wethink/demography#:~:text=Research%20conducted%20shows%20that%2C%20in,the%20current%20rate%20of%20prevalence.
30. Sullivan, P.F., et al., *Family history of schizophrenia and bipolar disorder as risk factors for autism.* Arch Gen Psychiatry, 2012. **69**(11): p. 1099-1103.
31. Dahan, S., *Surface Vision Transformers: Flexible Attention-Based Modelling of Biomedical Surfaces.* 2022.
32. Khayatzadeh-Mahani, A., et al., *Prioritizing barriers and solutions to improve employment for persons with developmental disabilities.* Disabil Rehabil, 2020. **42**(19): p. 2696-2706.
33. D'Souza, H. and A. Karmiloff-Smith, *Neurodevelopmental disorders.* Wiley Interdiscip Rev Cogn Sci, 2017. **8**(1-2).
34. Digital, N. *Autism Waiting Time Statistics - Quarter 1 To Quarter 4 2019-20 And Quarter 1 (April To June) 2020-21 - NHS Digital*. [cited 2022; Available from: https://digital.nhs.uk/data-andinformation/publications/statistical/autism-statistics/q1-april-to-june-2020-21#highlights.
35. Broder, M.S., et al., *Economic Burden of Neurologic Toxicities Associated with Treatment of Patients with Relapsed or Refractory Diffuse Large B-Cell Lymphoma in the United States.* Am Health Drug Benefits, 2020. **13**(5): p. 192-199.
36. Klinger, N.V. and S. Mittal, *Clinical efficacy of deep brain stimulation for the treatment of medically refractory epilepsy.* Clin Neurol Neurosurg, 2016. **140**: p. 11-25.
37. Nooyens, A.C.J., et al., *Sex Differences in Cognitive Functioning with Aging in the Netherlands.* Gerontology, 2022: p. 1-11.
38. Abdulkadir, M.B., W.B.R. Johnson, and R.M. Ibraheem, *Validity and accuracy of maternal tactile assessment for fever in under-five children in north central Nigeria: a cross-sectional study.* BMJ Open, 2014. **4**(10): p. e005776.
39. Rao, A., et al., *Predictive modelling using neuroimaging data in the presence of confounds.* Neuroimage, 2017. **150**: p. 23-49.
40. Snoek, L., S. Miletic, and H.S. Scholte, *How to control for confounds in decoding analyses of neuroimaging data.* Neuroimage, 2019. **184**: p. 741-760.
41. Abraham, A., et al., *Machine learning for neuroimaging with scikit-learn.* Front Neuroinform, 2014. **8**: p. 14.
42. Petch, J., S. Di, and W. Nelson, *Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology.* Can J Cardiol, 2022. **38**(2): p. 204-213.
43. Chang, L.J., et al., *Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference.* Cereb Cortex, 2013. **23**(3): p. 739-49.
44. Uddin, L.Q., et al., *Structure and Function of the Human Insula.* J Clin Neurophysiol, 2017. **34**(4): p. 300-306.
45. Specht, K. and P. Wigglesworth, *The functional and structural asymmetries of the superior temporal sulcus.* Scand J Psychol, 2018. **59**(1): p. 74-82.
46. Deen, B., et al., *Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus.* Cereb Cortex, 2015. **25**(11): p. 4596-609.
47. Bigler, E.D., et al., *Superior temporal gyrus, language function, and autism.* Dev Neuropsychol, 2007. **31**(2): p. 217-38.
48. Mullard, A., *The UK Biobank at 20.* Nat Rev Drug Discov, 2022.
49. Figueroa, C.A., et al., *Altered ability to access a clinically relevant control network in patients remitted from major depressive disorder.* Hum Brain Mapp, 2019. **40**(9): p. 2771-2786.
50. Auerbach, R.P., *RDoC and the developmental origins of psychiatric disorders: How did we get here and where are we going?* J Child Psychol Psychiatry, 2022. **63**(4): p. 377-380.
51. Seger, C.A., et al., *Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI.* Neuropsychologia, 2000. **38**(9): p. 1316-24.
52. Compton, R.J. and D.H. Weissman, *Hemispheric asymmetries in global-local perception: effects of individual differences in neuroticism.* Laterality, 2002. **7**(4): p. 333-50.

APPENDIX I

DESIGN STATEMENT

Deliverables	May	Jun	Jul	Aug
Literature Review				
Hand-engineer features				
Apply Deconfounding				
Model Training and Feature Selection				
Report				

Gantt chart detailing the project timeline and deliverables.

APPENDIX II

FURTHER INTERPRETATION OF RESULTS

Before deconfounding, the Lasso regressor closely follows the stacking model's performance compared with the other ensemble learning methods, despite ensemble methods usually being less prone to overfitting. However, Lasso regularised regression yields sparser solutions which may have discarded further redundant features, potentially explaining its unexpectedly high performance. After deconfounding, table 4 highlights how the stacking model still has the best predictive value in the context of fluid intelligence. More specifically, the performance is best when using half the number of features which could be due to redundant features being excluded. When performing further feature selection, the stacking predictive value lessens with a quarter of the original dataset, potentially reflecting that certain salient features were being overlooked.