Deep Learning vs. Classical Machine Learning: A Comparison of Methods for Fluid Intelligence Prediction

Luke Guerdan*[1], Peng Sun*[1], Connor Rowland[1], Logan Harrison[1], Zhicheng Tang[1], Nickolas Wergeles[1], and Yi Shang[1]

University of Missouri Columbia, Columbia MO, 65021, USA
{lmg4n8,ps793,carfzf,lkh6yb,zt253,wergelesn, shangy}@missouri.edu

Abstract. Predicting fluid intelligence based on T1-weighted magnetic resonance imaging (MRI) scans poses several challenges, including developing an adequate data representation of three dimensional voxel data, extracting predictive information from this data representation, and devising a model that is able to leverage the predictive information. We evaluate two strategies for prediction of fluid intelligence given structural MRI scans acquired through the Adolescent Brain Cognitive Development (ABCD) Study: deep learning models trained on raw imagery and classical machine learning models trained on extracted features. Our best-performing solution consists of a classical machine learning model trained on a combination of provided brain volume estimates and extracted features. Specifically, a Gradient Boosting Regressor (GBR) trained on a PCA-reduced feature space produced the best performance (train MSE = 66.29, validation MSE = 70.16), surpassing regression models trained on the provided volume data alone, and 2D/3D Convolutional Neural Networks trained on various representations of imagery data. Nonetheless, these results remain slightly better than a mean prediction, suggesting that neither approach is capturing a high degree of variance in the data.

Keywords: Image Processing· Neurocognitive Prediction · Machine Learning

1 Introduction

An ongoing challenge in neuroscience is relating brain structure to function, both at a neural scale and at the level of phenotypic expression. Though new neuroimaging approaches such as functional magnetic resonance imaging (fMRI) and diffusion tensor tractography have begun to shed light into this area, relating basic structural properties to complex behavioral expression remains difficult [1,2,3]. T1-weighted MRI is one neuroimaging method which has been used to relate brain structure to progression of autism, Alzheimer's, and Parkinson's

* denotes equal contribution

[4,5,6]. However, these abnormalities are often associated with gross differences in brain tissue, and the work employing structural MRI to discern more nuanced differences in regular brain function remains limited. Methodological advances in this area could prompt neuroscientific discoveries and objective assessment methods for normal and abnormal brain function.

One specific behavioral measure presumably tied to brain structure is fluid intelligence, especially fluid intelligence, which is central in abstract reasoning and problem solving [7]. Fluid intelligence has been linked to fronto-parietal connection properties [9], and a number of other brain characteristics [10]. Yet, no literature found by the authors directly links MRI-based structural information to fluid intelligence. A large corpus of data available through the Adolescent Brain Cognitive Development (ABCD) Study provides an unprecedented opportunity to investigate potential links between fluid intelligence and structural features of the brain. What's more, the ABCD Neurocognitive Prediction Challenge offers a structured context in which to develop pipelines for inferring fluid intelligence scores from scans acquired in the ABCD study. This work proposes a prediction pipeline addressing this challenge consisting of a GBR trained on ROI shape features. Results based on other regressors trained on extracted shape features, and results based on CNNs trained on representations of raw imagery, are also provided for comparison.

2 Related Work

A key issue to be addressed in a neurocognitive prediction context is the information representation, as MRI data natively occupies a three dimensional voxel space. One basic approach is to consider volume estimates (or other derived features) of brain regions, as the relative size and form of each region may contribute to cognitive function. Yet another approach is to examine patterns in the raw MRI imagery using computer vision methods. The information representation able to best capture variance in the data related to fluid intelligence may lead to the best results.

2.1 Raw Imagery

A recent trend in computer vision is replacing hand-engineered features with the unprocessed image to create an end-to-end prediction pipeline. This approach has proved particularly effective when combined with deep learning methods such as convolutional neural networks (CNNs), which have demonstrated the ability to extract salient features directly from the input [13]. However, most of these methods have been applied in the context of a two dimensional input image, as opposed to a 3D voxel space. One approach is to slice into the three dimensional space, however, this raises the issue of which slice is most discriminative, and may leave valuable information un-processed. A 2.5D CNN, which ingests slices from the axial, coronal, and sagittal planes, is one solution addressing this problem. This approach has proven useful for generating state-of-the-art

brain segmentation results [16]. Yet, this strategy also poses the challenge of the optimal slice to extract from each dimension. This motivates a 3D CNN which can ingest full imagery data from the entire voxel space. A 3D CNN has been employed for Alzheimers disease diagnostics, and demonstrated high predictive accuracy [5].

2.2 Derived Features

One direct method of representing brain structure is calculating features based on brain regions of interest, then training a machine learning model on these features [4,8,12]. For example, a deep autoencoder trained on brain region volume features predicted autism onset with high accuracy [4]. Another work successfully leveraged global and local shape features of tumor areas to predict benign vs. malignant tumor regrowth [12]. Specifically, this approach extracted global features such as elongation, compactness, volume, and surface area, as well as local characteristics calculated from Gaussian and mean curvatures of a constructed isosurface. Global shape features were calculated using the Insight Segmentation and Registration Toolkit (ITK) implementation [14]. Not only did this method achieve high accuracy, but it also conferred the advantage of revealing which features are most discriminative of tumor malignancy [12]. The interpretability of deep learning methods, especially those leveraging raw imagery data, is less direct.

3 Methods

Approaches for predicting fluid intelligence based on MRI imagery based on both (a) raw imagery, and (b) derived features were developed and tested. We predicted that deep learning methods ingesting raw MRI imagery would ultimately have more representational power and would demonstrate better results. However, due to the complexity of the raw MRI imagery, the deep learning models we tested did not yield better results with this dataset. Though classical machine learning models trained on derived features provided lower mean squared error, we still provide details regarding the deep learning methods developed for comparison.

3.1 Deep Learning Methods

In order to extract shape information for each brain region, we convolved across voxels in three dimensions using a 3D CNN. This is motivated by recent work indicating CNNs can extract meaningful shape patterns in image classification [15] and object detection [11] contexts. Our pipeline for deep learning based on raw imagery is shown in Fig. 1. We adapted a 3D-ResNet [18], which was originally used for video classification, to integrate information across slices. Two specific adaptions were required. (1) Since the third dimension is temporal correlation in a video classification task, this was changed to spatial correlation

over the input shape. (2) The final layer was updated to consist of a single node outputting the predicted fluid intelligence score.

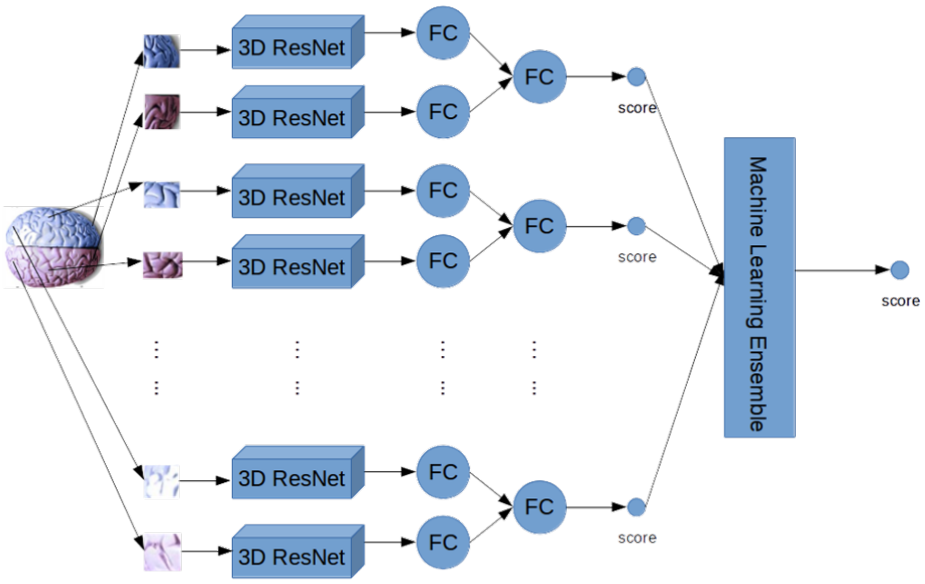

Fig. 1. 3D CNN for prediction of fluid intelligence score for each brain region. Each hemisphere is fed into a 3D ResNet separately and then results are stacked into one connected layer together. The score predicted from each region (both left and right hemispheres) is then aggregated using a classical machine learning ensemble method to generate a final score.

As shown in Fig. 1, feature extraction from each provided region of segmented grey matter [17] was first conducted using the previously described 3D CNN. The proceeding fully connected layers were organized hierarchically, with the first layer matching extracted features across hemispheres. A model was trained for each brain region (including left and right hemisphere), then predicted scores were ensembled using a regression model. A preliminary assessment of regression models including AdaBoost, Ridge regression, Random Forest, and SVR showed that AdaBoost provided the best performance, which was used in the reported experiments. This was performed for both grey matter and non-grey matter in the region of interest.

3.2 Shape-Based Machine Learning Methods

We also developed a preprocessing pipeline which ingests the skull stripped and raw MRI images provided for each participant, and extracts global shape features based on isolated ROIs, full grey matter, and full non-grey matter as calculated

by [17]. Version 4.13.1 of ITK was used to extract (a) volume, (b) elongation, (c) surface area, (d) roundness, and (e) flatness from each of the 100 largest regions provided in the segmented grey matter [14].

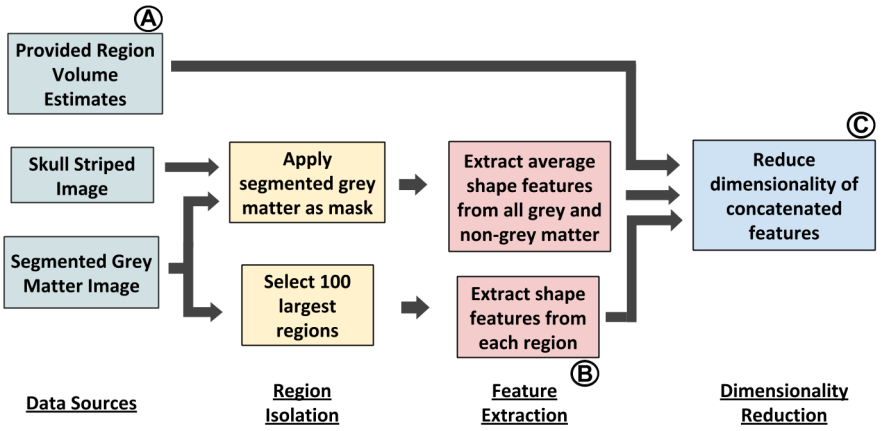

Fig. 2. Pipeline for extracting and aggregating features used in regressor training (classical machine learning models). **A.** shows the dataset for provided estimates alone,**B.** shows the dataset consisting of only ROIs, while **C.** shows the dataset consisting of all features.

The ordering of these 100 regions was fixed based on the first training participant and kept constant for the training, validation, and test data. ITK returns extracted features for each of the non-contiguous segments of the isolated ROI, and we used the first (largest) segment returned. Shape features from the overall grey and non-grey matter were extracted according to the same process above, except that each shape feature was averaged for all segments returned. Non-grey matter was isolated by applying the grey matter as a mask to the overall skull-stripped image for each participant (shown in Fig. 3). This non-grey matter was included to provide some proxy of white matter, which has been shown to be a relevant factor for fluid intelligence in the literature. Including this information empirically showed a slight performance improvement in validation loss.

The feature extraction process resulted in a 500 dimensional feature space generated by the ROI shape extraction, and a 10 dimensional feature space generated by global feature extraction. Each of these features were then combined with the provided volume estimates to give a final dataset for training. Since this dataset occupied a high-dimensional space, dimensionality reduction (by PCA) was used to project the data to a lower dimensional subspace. A series of regression models were then trained on the PCA-reduced final dataset (Fig. 2.C), as well as the given volume data alone (Fig. 2.A), and the extracted ROI features alone (Fig. 2.B). The specific models trained included lasso, ridge, support vector, gradient boosting, and AdaBoost regressors. A grid search was performed

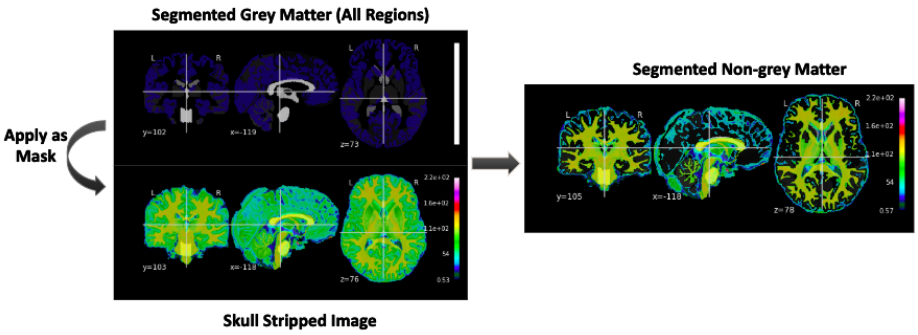

Fig. 3. Process of isolating grey and non-grey matter for feature extraction using masking.

to determine the optimal parameters for training on all datasets, and PCA was included in this parameter grid search for the (Fig. 2.C) dataset. Models were evaluated by the MSE error metric, which defines the error between the predicted labels $\hat{Y}$ and actual labels Y over n samples as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{1}$$

The train, validation, test split aligned with the data provided in the ABCD Neurocognitive Prediction Challenge, with 3739 samples used for training, 415 samples used for validation, and 4402 samples left out for testing. In order to establish a baseline estimate of the predictive value of each model, the error of predicting the mean was first established by calculating the average score from the training dataset and calculating the MSE between this mean and each label in the training and validation datasets. A series of regressors were then trained on each of the three datasets described above, including Ridge and Lasso regression, SVR, GBR, and AdaBoost.

4 Results

Table 1 reports training and validation error obtained from training the five algorithms above on dataset A, B, and C (Fig. 2), as well as results from the best-performing 3D CNN models. Fig. 4 shows a visual comparison of validation loss provided in the table. Overall, the GBR demonstrated the highest predictive value for each of the three extracted feature datasets, with the best performing model being a GBR trained on the PCA-reduced full features. This reached a $MSE(train) = 66.29$ and $MSE(test) = 70.16$. The optimal configuration for this model consisted of 95 PCA components and 45 learners in the GBT. Compared with 3D CNN models, as shown in Table 1, the GBT outperformed 3D CNN models by 1.14 MSE in validation error.

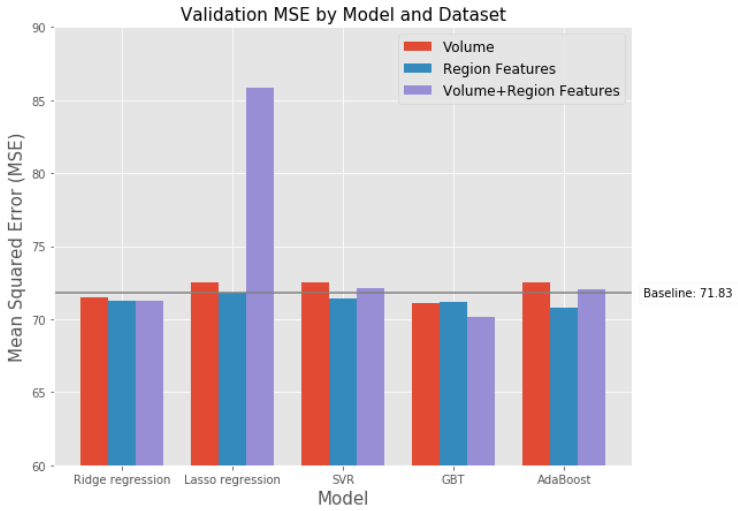

Fig. 4. Validation MSE for each of the five models trained on the three datasets. The horizontal bar indicates the MSE obtained by predicting the mean of the training dataset.

Dataset	Method	Train (MSE)	Val.(MSE)
	Mean prediction	**85.84**	**71.83**
Volume (A)	Ridge reg.	82.64	71.52
	Lasso reg.	81.62	72.53
	SVR	83.06	72.51
	GBR	80.16	71.1
	AdaBoost	79.60	72.49
Extracted Features (B)			
	Ridge reg.	78.28	71.22
	Lasso reg.	85.84	71.83
	SVR	77.88	71.43
	GBR	73.50	71.21
	AdaBoost	77.85	70.79
Volume + Extracted Features (C)			
	Ridge reg.	81.84	71.22
	Lasso reg.	71.83	85.84
	SVR	72.22	72.16
	GBR	**66.29**	**70.16**
	AdaBoost	80.10	72.03
White Matter Imagrey	3D CNN	85.43	71.64
Grey Matter Imagery	3D CNN	85.52	71.75
White+Grey Model Imagery	3D CNN ensemble	84.55	71.30

Table 1. Train and validation MSE obtained from all methods and datasets.

5 Discussion

Overall, it proved difficult to find a model which could capture a large degree of variance in the data. The best validation score obtained indicated a R2 value of .03, which captures a modest 3 percent of the variance in the data. Most of the derived features and models demonstrate menial incremental improvements as opposed to qualitative advances in neurocognitive prediction. Though we predicted that deep learning would provide a robust predictive strategy, this was not the case. We suspect the reason is that (1) excessive variability in the features extracted between subjects or (2) the input images are too complex for a CNN to learn relevant features. Though the ABCD dataset contains a large amount of data for a neuroscience context, this quantity still falls very short of the enormous corpuses typically used in other image processing domains. Therefore, additional data could help aid prediction. However, it is possible that given the correct way of representing the data, better results could be obtained. Models such as autoencoder could also provide predictive value, as they have in similar contexts [4].

It is also possible that structural features alone do not contain enough information related to fluid intelligence to be useful in prediction contexts. Other information such as cortical connectivity and functional activation may be required to capture an accurate assessment of ones fluid intelligence. Combining multi-modal neuroimaging datasets presents yet another representational problem, but could prove fruitful in bolstering model performance.

References

1. Le Bihan, Denis, et al. "Diffusion tensor imaging: concepts and applications." Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 13.4 (2001): 534-546.
2. Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., Shulman, G. L. (2001). A default mode of brain function. Proceedings of the National Academy of Sciences, 98(2), 676-682.
3. Bullmore, E., Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience, 10(3), 186.
4. Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., ... Collins, D. L. (2017). Early brain development in infants at high risk for autism spectrum disorder. Nature, 542(7641), 348.
5. Hosseini-Asl, E., Keynton, R., El-Baz, A. (2016, September). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 126-130). IEEE.
6. Morales, D. A., Vives-Gilabert, Y., Gmez-Ansn, B., Bengoetxea, E., Larraaga, P., Bielza, C., ... Delfino, M. (2013). Predicting dementia development in Parkinson's disease using Bayesian network classifiers. Psychiatry Research: NeuroImaging, 213(2), 92-98.
7. Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. Intelligence, 28(2), 121-143.

8. Sun, Peng, et al. "Ada-automatic detection of alcohol usage for mobile ambulatory assessment." 2016 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, 2016.
9. Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J. H., Lee, S., Kim, K. (2006). Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. Neuroimage, 29(2), 578-586.
10. Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., Alkire, M. T. (2004). Structural brain variation and general intelligence. Neuroimage, 23(1), 425-433.
11. Liu, Yang, et al. "Performance comparison of deep learning techniques for recognizing birds in aerial images." 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE, 2018.
12. Ismail, M., Hill, V., Statsevych, V., Huang, R., Prasanna, P., Correa, R., ... Madabhushi, A. (2018). Shape Features of the Lesion Habitat to Differentiate Brain Tumor Progression from Pseudoprogression on Routine Multiparametric MRI: A Multisite Study. American Journal of Neuroradiology, 39(12), 2187-2193.
13. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
14. Kim, Y. (2001). Insight Segmentation and Registration Toolkit. The National Library of Medicine, Washington, DC.
15. Chen, G., Sun, P. and Shang, Y., 2017, November. Automatic Fish Classification System Using Deep Learning. In 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 24-29). IEEE.
16. Kushibar, K., Valverde, S., Gonzlez-Vill, S., Bernal, J., Cabezas, M., Oliver, A., Llad, X. (2018). Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. Medical image analysis, 48, 177-186.
17. Pfefferbaum, A., Kwon, D., Brumback, T., Thompson, W. K., Cummins, K., Tapert, S. F., ... Sullivan, E. V.(2017). Altered brain developmental trajectories in adolescents after initiating drinking. American journal of psychiatry, 175(4), 370-380.
18. Hara, Kensho, Hirokatsu Kataoka, and Yutaka Satoh. "Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018.