RESEARCH ARTICLE OPEN ACCESS

The Shape of the Brain's Connections Is Predictive of Cognitive Performance: An Explainable Machine Learning Study

Yui Lo[1,2,3] | Yuqian Chen[1,2] | Dongnan Liu[3] | Wan Liu[4] | Leo Zekelman[2,5] | Jarrett Rushmore[6,7] | Fan Zhang[8] | Yogesh Rathi[1,2] | Nikos Makris[1,6] | Alexandra J. Golby[1,2] | Weidong Cai[3] | Lauren J. O'Donnell[1,2,9]

[1]Harvard Medical School, Boston, Massachusetts, USA | [2]Brigham and Women's Hospital, Boston, Massachusetts, USA | [3]The University of Sydney, Sydney, Australia | [4]Beijing Institute of Technology, Beijing, China | [5]Harvard University, Boston, Massachusetts, USA | [6]Massachusetts General Hospital, Boston, Massachusetts, USA | [7]Boston University, Boston, Massachusetts, USA | [8]University of Electronic Science and Technology of China, Chengdu, China | [9]Harvard-MIT Health Sciences and Technology, Cambridge, Massachusetts, USA

Correspondence: Fan Zhang (zhangfanmark@gmail.com) | Weidong Cai (tom.cai@sydney.edu.au) | Lauren J. O'Donnell (odonnell@bwh.harvard.edu)

ABSTRACT

The shape of the brain's white matter connections is relatively unexplored in diffusion magnetic resonance imaging (dMRI) tractography analysis. While it is known that tract shape varies in populations and across the human lifespan, it is unknown if the variability in dMRI tractography-derived shape may relate to the brain's functional variability across individuals. This work explores the potential of leveraging tractography fiber cluster shape measures to predict subject-specific cognitive performance. We implement two machine learning models (1D-CNN and Least Absolute Shrinkage and Selection Operator [LASSO]) to predict individual cognitive performance scores. We study a large-scale database from the Human Connectome Project Young Adult study ($n = 1065$). We apply an atlas-based fiber cluster parcellation (953 fiber clusters) to the dMRI tractography of each individual. We compute 15 shape, microstructure, and connectivity features for each fiber cluster. Using these features as input, we train a total of 210 models (using fivefold cross-validation) to predict 7 different NIH Toolbox cognitive performance assessments. We apply an explainable AI technique, SHapley Additive exPlanations (SHAP), to assess the importance of each fiber cluster for prediction. Our results demonstrate that fiber cluster shape measures are predictive of individual cognitive performance. The studied shape measures, such as irregularity, diameter, total surface area, volume, and branch volume, are generally as effective for prediction as traditional microstructure and connectivity measures. The 1D-CNN model generally outperforms the LASSO method for prediction. Further interpretation and analysis using SHAP values from the 1D-CNN suggest that fiber clusters with features highly predictive of cognitive ability are widespread throughout the brain, including fiber clusters from the superficial association, deep association, cerebellar, striatal, and projection pathways. This study demonstrates the strong potential of shape descriptors to enhance the study of the brain's white matter and its relationship to cognitive function.

1 | Introduction

Diffusion magnetic resonance imaging (dMRI) tractography is an in vivo imaging method to map white matter connections based on the water diffusion in brain tissue (Basser et al. 2000). dMRI tractography produces sequences of points called streamlines that estimate the course of white matter connections. Streamlines can be organized into anatomical fiber tracts or finely parcellated into fiber clusters, enabling quantitative measures of white matter microstructure, connectivity, and shape (Zhang, Daducci et al. 2022). These measures allow the study of the brain's white matter in health and disease and its relationship to non-imaging phenotypes, such as individual cognitive performance (Forkel et al. 2022). Recent studies have demonstrated the prediction of non-imaging phenotypes from microstructure and connectivity measures at the fiber tract and fiber cluster levels (Chamberland et al. 2021; Chen, Zekelman et al. 2024; Xiao et al. 2021; Zekelman et al. 2022). These studies have employed regression or deep learning models to output a subject-specific non-imaging phenotype score. However, these studies mainly investigated traditional dMRI tractography measures, such as the number of streamlines (NoS), fractional anisotropy (FA), or mean diffusivity (MD). The NoS relates to the geometry of the connection but is usually thought of as a proxy for its connectivity or connection "strength" (Zhang, Daducci et al. 2022), while the FA and MD microstructure measures quantify the anisotropy and magnitude of water diffusion and are sensitive to a variety of tissue properties (Beaulieu 2009; Jones, Knösche, and Turner 2013).

While these microstructure and connectivity features are highly informative, they ignore many potentially important morphometric or shape characteristics of the brain's connections. Recently, shape measures, including length, area, volume, and other metrics, were applied to study the morphometry of the brain's association fiber tracts (Yeh 2020). Other research has shown that shape measures vary across the lifespan (Lebel et al. 2012; Schilling, Archer et al. 2023; Schilling, Chad et al. 2023). While these studies have demonstrated that

measures of fiber tract shape have both population and lifespan variability, it is unknown to what extent the dMRI tractography shape variability may relate to the brain's functional variability across individuals. Our recent preliminary work has shown that fiber cluster shape features have the potential to predict individual language functional performance (Lo, Chen et al. 2024). However, it is not yet known if dMRI tractography shape features are predictive of different cognitive abilities, which shape features may be the most predictive, and how their predictive performance may compare to more traditional features such as NoS and FA. The answers to these questions may help guide further research into the shape of the brain's connections.

To begin to address these questions, the main contribution of this work is to explore the potential of a comprehensive set of fiber cluster shape measures for the prediction of multiple domains of individual cognitive performance. We propose a computationally intensive data-driven strategy of training many (210) machine learning models to predict individual cognitive performance using multiple input microstructure, connectivity, and shape measures. By comparing the models' performance, we aim to achieve a broad, data-driven assessment of the potential of shape for studying the brain's white matter connections and their relationship to human brain functional performance.

2 | Materials and Methods

2.1 | Overview

Our overall strategy is as follows (Figure 1). We first perform a fine parcellation of the white matter into fiber clusters (Section 2.2.1), which have been shown to have improved power to predict human traits in comparison with traditional connectome matrices (Liu, Li, and Dunson 2023). We then calculate shape measures (12 features described below in Section 2.3.2), microstructure features (FA and MD), and NoS features for each fiber cluster for all subjects. Next, we use these features as inputs to train machine learning models
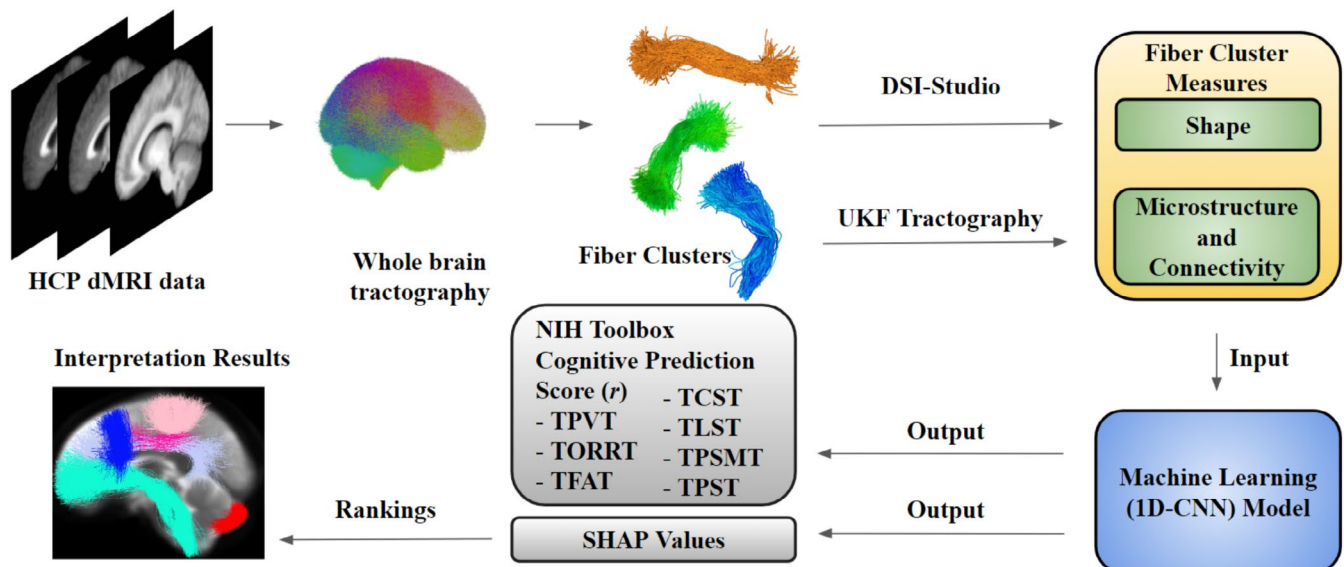

FIGURE 1 | The overall data-driven pipeline of our work.

TABLE 1 | Overview of NIH Toolbox cognitive assessments used in this study.

Category	Assessment	Abbreviation	Description	Citation
Language	NIH Toolbox Picture Vocabulary Test	TPVT	Tests the ability to select the picture that corresponds to a spoken word	(Gershon et al. 2014)
	NIH Toolbox Oral Reading Recognition Test	TORRT	Assesses the ability to pronounce individual printed words and identify letters	(Zelazo et al. 2013)
Executive function and attention	NIH Toolbox Flanker Inhibitory Control and Attention Test	TFAT	Assesses the ability to focus on a particular stimulus while inhibiting attention to the stimuli flanking it	
	NIH Toolbox Dimensional Change Card Sort Test	TCST	Assesses the flexibility to match a series of picture pairs to a target picture	(Tulsky et al. 2014)
Working and episodic memory	NIH Toolbox List Sorting Working Memory Test	TLST	Assesses working memory to recall and sequence different stimuli that are presented visually and via audio	(Bauer et al. 2013)
	The NIH Toolbox Picture Sequence Memory Test	TPSMT	Assesses episodic memory to remember a sequence of pictures of objects and activities	
Processing speed	NIH Toolbox Pattern Comparison Processing Speed Test	TPST	Measures the ability to quickly determine whether two stimuli are the same or different	(Carlozzi et al. 2015)

(CNN and Least Absolute Shrinkage and Selection Operator [LASSO]) that are known to successfully predict non-imaging phenotypes from tractography data (Chen et al. 2022; Cui and Gong 2018; Feng et al. 2022; He et al. 2022; Liu, Chen et al. 2023; Lo, Chen et al. 2024; Wen et al. 2019). We train the models (Section 2.4) to predict individual cognitive performance, where we use a testbed of 7 NIH Toolbox assessments (described below in Section 2.2.2). Finally, the predictive performance of the models is compared (Section 2.5) to assess the utility of the input measures for predicting cognitive performance. We implement an explainable module (Section 2.6) to describe the contribution of each fiber cluster measure to the prediction, and we provide example visualizations of the predictive white matter anatomy.

2.2 | Study Material

2.2.1 | Dataset, Tractography, and Fiber Clustering Parcellation

In this work, we study the brain connections of 1065 healthy young adults (YAs; 575 females and 490 males, mean age 28.7 years) from the Human Connectome Project minimally preprocessed YA dataset (HCP-YA) (Van Essen et al. 2012; Van Essen et al. 2013). We used pre-computed HCP-YA tractography as employed in recent work (Chen et al. 2023; Zekelman et al. 2022). Whole brain tractography was performed with a multi-tensor unscented Kalman filter (UKF) tractography method (Reddy and Rathi 2016)[1] that is highly consistent across the human lifespan, across test–retest scans, across disease states, and across different acquisitions (Zhang, Wu et al. 2018; Zhang et al. 2019). UKF is effective for reconstructing anatomical somatopy (He et al. 2023). The UKF method estimates a tissue microstructure model during fiber tracking, leveraging prior information from the previous tracking step to increase model fitting stability. UKF enables the estimation of tract-specific microstructural measures using the first tensor, which models the traced tract. Each subject's whole brain tractography was automatically parcellated into fiber clusters using whitematteranalysis (Zhang, Wu, Norton et al. 2018). This atlas-based machine learning method is consistent across the human lifespan, across test–retest scans, across disease states including brain tumors, and across different acquisitions (Zhang, Wu, Norton et al. 2018)[2] with high test–retest reproducibility (Zhang et al. 2019). Fiber clusters are known to provide a compact vector representation of the connectome with improved power to predict human traits (Liu, Li, and Dunson 2023; Zhang, Savadjiev et al. 2018), enabling a variety of downstream analyses (Chen et al. 2023; Gabusi et al. 2024; Xue et al. 2024; Zhang, Wu, Ning et al. 2018). The whitematteranalysis method parcellates robustly by using a spectral embedding of streamlines, a machine-learning technique that takes into account the variability across subjects (O'Donnell and Westin 2007). Fiber clusters were defined in the whitematteranalysis package using the O'Donnell Research Group (ORG) fiber cluster atlas (Zhang, Wu, Norton et al. 2018). The ORG atlas was anatomically curated to organize fiber clusters into anatomically labeled fiber tracts and to enable the automatic removal of anatomically inaccurate (false positive) fiber clusters. For each subject in the current study, the whitematteranalysis processing provides 953 fiber clusters,

anatomically categorized into 58 deep white matter anatomical tracts, including association, cerebellar, commissural, and projection tracts, and 16 superficial tract categories based on location within different brain lobes.

2.2.2 | NIH Toolbox Cognitive Performance Assessments

In this work, we investigate the potential of shape measures for predicting non-imaging phenotypes. As a testbed for prediction, we choose seven non-imaging cognitive phenotypes that span the domains of language, executive function and attention, working and episodic memory, and processing speed (Table 1). These phenotypes are from the NIH Toolbox, a standard battery for neurobehavioral measurement (Hodes et al. 2013). Specifically, we use the NIH Toolbox age-adjusted cognitive performance assessments provided by HCP-YA (Weintraub et al. 2013). These assessments (Table 1) are the NIH Toolbox Picture Vocabulary Test, the NIH Toolbox Oral Reading Recognition Test, the NIH Toolbox Flanker Inhibitory Control and Attention Test, the NIH Toolbox Dimensional Change Card Sort Test, the NIH Toolbox List Sorting Working Memory Test, the NIH Toolbox Picture Sequence Memory Test, and the NIH Toolbox Pattern Comparison Processing Speed Test.

2.3 | Fiber Cluster Measures

2.3.1 | Microstructure and Connectivity Measures

We study 3 widely used FA, MD, and NoS measures, which are calculated for each of the 953 fiber clusters from all subjects. These popular measures are informative in the study of the relationship between the brain's fiber tracts and cognition (Chen, Zekelman et al. 2024; Forkel et al. 2022; Zekelman et al. 2022). FA measures how much a tensor deviates from a sphere, quantifying the anisotropy of water molecule diffusion. MD reflects the total amount of water diffusion within the tissue, providing an average measure of the diffusion in all directions. NoS can be considered a geometry or shape measure but is popularly considered a "connectivity" measure and is indirectly related to structural connectivity "strength" (Qi et al. 2015; Zhang, Daducci et al. 2022).

2.3.2 | Shape Measures

We study 12 fiber cluster shape measures that are considered to provide a detailed and comprehensive shape analysis of tractography (Yeh 2020). These measures include length, span, volume, diameter, curl, elongation, trunk volume, branch volume, total surface area, total area of end regions, total radius of end regions, and irregularity (Figure 2). These shape measures are computed for all 953 fiber clusters from all subjects by applying the software package DSIStudio (Yeh 2020)[3].

Full definitions of the shape measures can be seen in (Yeh 2020), and we include a brief description of each measure here. Length is the average streamline length of the fiber cluster in mm. Span is the distance between the two ends of the fiber cluster in mm. Volume is defined as the volume of
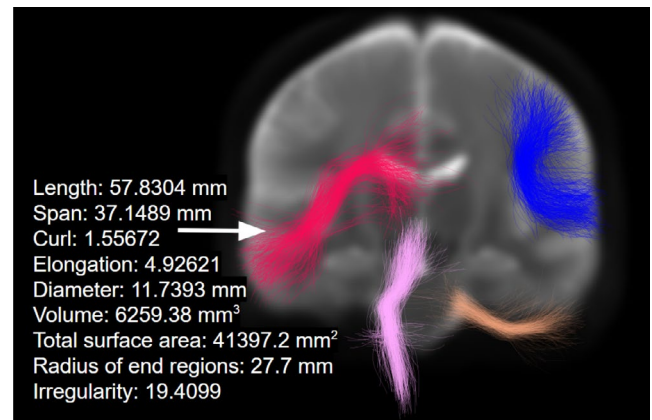

FIGURE 2 | Four examples of individual white matter connections (fiber clusters) extracted from the entire white matter of the human brain using a fiber clustering approach. Example shape descriptors are shown for the red fiber cluster.

voxels occupied by the fiber cluster in mm^3, where DSIStudio creates a synthetic volume of isotropic voxels for use in computations. Diameter is estimated from the length and volume shape measures, under the assumption that the fiber cluster takes the form of a cylinder; the diameter is considered to represent the average fiber cluster diameter in mm. Curl is defined as the length divided by the span to approximate the overall curve (note this is also known as the u-ratio in other literature, e.g., Nie et al. 2024). Elongation is defined as the fiber cluster length divided by the diameter. The trunk volume is the volume of the streamlines connecting the fiber cluster's two largest end regions. The branch volume is the remaining volume (the difference between the total and trunk volume). The surface area is estimated by counting the outermost or surface voxels occupied by the fiber cluster, then multiplying by the area of one side of a voxel in mm^2. The total area of end regions is computed as above using the surface voxels of the cluster endpoints, which are also used in estimating the total radius of end regions using a circular model. Finally, irregularity is computed by dividing the fiber cluster surface area by the surface area of a cylinder with the same diameter and length; thus, irregularity quantifies how different the shape is from an idealized cylinder.

2.4 | Machine Learning Prediction

To assess whether the fiber cluster shape measures help predict individual cognitive performance, we employ two machine learning prediction methods: one deep learning method and one traditional method.

First, we implement a deep learning method using a 1D CNN model. This approach has previously successfully predicted non-imaging phenotypes from tractography data (He et al. 2022; Liu, Chen et al. 2023; Lo, Chen et al. 2024). Specifically, the input of each model is one of the 15 shape, microstructure, or connectivity measures measured from all 953 fiber clusters. The model output is a predicted cognitive performance score (one of the seven described above). In technical details, the model's architecture is designed as follows.

The model consists of a feature extractor and a classifier. The model is trained using mean squared error loss by comparing its output with the ground truth. The feature extractor contains three convolutional blocks, each comprising a 1D convolutional layer (kernel size = 5, kernel number = 64, stride = 1), a batch normalization layer with a momentum of 0.5, an ReLU activation layer, and a dropout layer for regularization. The classifier comprises two fully connected layers; the first layer reduces the input dimension from 512 to 128, followed by batch normalization with a momentum of 0.5, an ReLU activation layer, and dropout. The final layer maps the output to the predicted score of the cognitive performance. This deep multilayered architectural design aims to effectively learn from the input measure to improve the overall prediction of the cognitive performance score.

For comparison, we implement LASSO regression (Feng et al. 2022; Tibshirani 2018), a conventional machine learning method that leverages variable selection and regularization. LASSO regression has been used successfully to predict cognitive performance from brain connectome and fiber tract data (Chen et al. 2022; Cui and Gong 2018; Feng et al. 2022; Wen et al. 2019).

Overall, in this project, a model is trained to predict each cognitive assessment from each input feature. This results in a total of 210 machine learning models were trained, including the 105 (15 input features × 7 cognitive assessments) 1D-CNN models and the 105 (15 × 7) LASSO models. This approach ensures that each model focuses on a specific input-prediction pair, reducing the complexity of training. This experimental design enables the comparison of prediction performance across all input features. Fivefold cross-validation was used for training each model (see Section 2.7 for more details).

2.5 | Statistical Analysis

We conduct a statistical analysis of the prediction performance results. The correlation coefficient (Pearson's r) is not normally distributed. Therefore, we transform the results from all five folds to z-scores using the widely applied Fisher's r-to-z transformation (Chen, Zekelman et al. 2024; Keller et al. 2011; Shen et al. 2015; Tobyne et al. 2018). To compare the performance of different models, we employ a one-way repeated measures analysis of variance (ANOVA) of the z-transformed correlation coefficients, followed by post hoc pairwise comparisons using paired t-tests between each of the different models.

2.6 | Interpretation With Explainable AI

To provide potential insight into how the shape, microstructure, and connectivity features inform the prediction of individual cognitive performance, we implement a widely used explainable AI approach, SHapley Additive exPlanations (SHAP). SHAP assigns each feature an importance value for a particular prediction (Lundberg and Lee 2017). The computation of SHAP values is based on a coalition game theory approach, providing a clear and interpretable explanation to improve the trustworthiness of the interpreted model. The

SHAP approach has been widely applied for the interpretation of brain features that affect the prediction of individual intelligence, memory, language (Azevedo et al. 2019), and brain age (Ballester et al. 2023; Scheda and Diciotti 2022; Sun et al. 2022). In this study, the objective of leveraging SHAP interpretation is to understand the model's insights into the brain's white matter connections and how their microstructure, connectivity, and shape features may relate to cognitive performance.

To interpret the contribution of each of the 953 fiber clusters, we evaluate which fiber clusters most impact the prediction results. To compute this, for each of the seven NIH toolbox assessment prediction models, we calculate each cluster's mean SHAP value across the five folds. Then, we rank the clusters for each model according to SHAP value. Finally, to assess the overall impact on prediction results across models, we average across the seven different rankings to get a mean ranking for each cluster. We sort clusters according to mean rankings in ascending order to evaluate the most impactful fiber clusters for the overall prediction of cognitive assessment scores.

2.7 | Implementation Details

We optimize our 1D-CNN model with the Stochastic Gradient Descent (SGD) algorithm (Zou and Hastie 2005). The SGD optimizer is instantiated with a learning rate of 0.01 and trained on 200 epochs with batch sizes of 8, and the number of workers is set to 10. All experiments in this work were performed on an NVIDIA RTX A5000 GPU using PyTorch 1.8.1 (Paszke et al. 2019) across a fivefold cross-validation to ensure reproducibility and consistency in model training and evaluation. We use the sklearn python package (Pedregosa et al. 2011) for the LASSO regression model with alpha set to 1.

All 105 1D-CNN model training experiments were run with the same fine-tuned hyperparameters. All 105 1D-CNN and 105 LASSO model training experiments were run with the same data split of 80% training data and 20% testing data across the five folds.

3 | Results

3.1 | Evaluation Metric

We employ the Pearson correlation coefficient (Pearson's r) (Sedgwick 2012) to evaluate model performance. This metric is widely applied in neurocognitive performance prediction (Chen et al. 2020; Feng et al. 2022; Gong, Beckmann, and Smith 2021; Jeong et al. 2021; Kim et al. 2021; Rasero et al. 2021; Tian and Zalesky 2021; Wu et al. 2023; Xue et al. 2024). Pearson's r measures the strength and direction (positive or negative) of the linear association between two variables.

3.2 | Prediction Performance Results

Table 2 gives the model performance (r) for the 105 (15 × 7) 1D-CNN models trained using the 15 input features (microstructure,

TABLE 2 | Pearson correlation coefficients (*r*) of the 1DCNN for predicting seven NIH Toolbox cognitive assessments. Values are the mean and standard deviation across fivefold cross-validation. The best result for each column is bolded. Results that are not significantly different from the best result in each column are in italics.

	Language		Executive function		Memory		Processing speed	Average *r*
	TPVT	TORRT	TFAT	TCST	TLST	TPSMT	TPST	
Microstructure								
FA	*0.314±0.08*	*0.281±0.06*	*0.132±0.03*	*0.053±0.03*	**0.239±0.03**	*0.128±0.03*	*0.135±0.03*	*0.183±0.1*
MD	*0.263±0.02*	*0.207±0.04*	*0.181±0.04*	*0.076±0.05*	*0.209±0.06*	**0.147±0.04**	*0.134±0.08*	*0.174±0.07*
Connectivity								
NoS	*0.328±0.07*	**0.294±0.05**	*0.179±0.05*	**0.148±0.07**	*0.229±0.07*	*0.096±0.04*	*0.116±0.04*	*0.199±0.09*
Shape								
Irregularity	*0.312±0.06*	*0.281±0.07*	**0.183±0.06**	*0.103±0.04*	*0.224±0.04*	*0.115±0.03*	*0.177±0.05*	**0.2±0.08**
Diameter	**0.335±0.06**	*0.289±0.06*	*0.166±0.06*	*0.119±0.06*	*0.199±0.02*	*0.09±0.03*	*0.173±0.03*	*0.196±0.09*
Total surface area	*0.318±0.05*	*0.277±0.07*	*0.15±0.05*	*0.109±0.02*	*0.192±0.03*	*0.106±0.04*	**0.188±0.05**	*0.191±0.09*
Branch volume	*0.316±0.09*	*0.283±0.07*	*0.134±0.05*	*0.126±0.05*	*0.209±0.03*	*0.1±0.01*	*0.155±0.06*	*0.189±0.09*
Volume	*0.325±0.07*	*0.253±0.09*	*0.122±0.05*	*0.126±0.04*	*0.215±0.03*	*0.102±0.03*	*0.175±0.06*	*0.188±0.09*
Surface area of end regions	*0.34±0.06*	*0.276±0.05*	*0.13±0.03*	*0.082±0.06*	*0.227±0.01*	*0.072±0.02*	*0.139±0.02*	*0.181±0.11*
Elongation	*0.245±0.05*	*0.212±0.03*	*0.157±0.06*	*0.143±0.04*	*0.192±0.06*	*0.054±0.04*	*0.159±0.03*	*0.166±0.07*
Radius of end regions	*0.229±0.05*	*0.24±0.04*	*0.139±0.03*	*0.11±0.04*	*0.155±0.04*	*0.109±0.05*	*0.162±0.05*	*0.164±0.06*
Curl	*0.189±0.08*	*0.208±0.10*	*0.139±0.06*	*0.067±0.03*	*0.153±0.06*	*0.115±0.05*	*0.146±0.10*	*0.145±0.05*
Span	*0.253±0.04*	*0.204±0.07*	*0.124±0.04*	*0.084±0.03*	*0.176±0.06*	*0.057±0.04*	*0.091±0.05*	*0.141±0.08*
Length	*0.214±0.07*	*0.185±0.07*	*0.087±0.05*	*0.093±0.04*	*0.131±0.08*	*0.071±0.04*	*0.092±0.04*	*0.125±0.06*
Trunk volume	*0.126±0.072*	*0.104±0.08*	*0.076±0.05*	*0.074±0.07*	*0.083±0.05*	*0.095±0.04*	*0.108±0.05*	*0.095±0.02*

connectivity, and shape) to predict the 7 output NIH toolbox assessment scores. The top-performing model in each column is shown in bold and represents the highest r for predicting a particular NIH toolbox assessment score. Other models with statistically equivalent performance to the top-performing model are shown in italics.

Overall, the results shown in Table 2 suggest that many shape features are as predictive as the widely used microstructure and connectivity features. Many features exhibit statistically equivalently high performance (in bold and italics), as shown in each column of Table 2. On average (rightmost column of Table 2), the highest-performing features include microstructure (FA and MD), connectivity (NoS), and the top-performing shape measures of irregularity, diameter, total surface area, branch volume, volume, surface area of end regions, and elongation. These features are similarly predictive overall on average across the seven prediction tasks. In fact, the overall highest-performing measure is a shape measure, Irregularity (though its performance is not significantly different from the other highest-performing measures). While the highest-performing shape measures have similar performance to the microstructure and connectivity features, it can be noted that several shape measures do have lower performance on average in this experiment (radius of end regions, curl, span, length, and trunk volume). In this experiment, the least informative feature is trunk volume.

For a complementary assessment of the utility of each feature to predict subject-specific cognitive performance, Table 3 summarizes the prediction performance (r) across the LASSO and CNN models. Note that for the 1D-CNN model, these values are the same as the values in the rightmost column of Table 2. Also, more details on the LASSO results are provided in Supplementary Table S1. Overall, the 1D-CNN method generally outperforms the LASSO method. However, the LASSO method has better worst-case performance and provides improved prediction performance for the lower-performing shape measures (curl, span, length, trunk volume), suggesting their potential utility. The statistical analysis (repeated measures ANOVA) of the LASSO results finds no significant performance difference across the models trained with the 15 features ($p = 0.37$), suggesting that all investigated microstructure, connectivity, and shape features have utility for the study of the brain.

3.3 | Explainable Interpretation

In this section, we provide potentially informative visualizations of the brain fiber clusters whose input features are most predictive of individual cognition. We focus on the interpretation of the best-performing models ($r > 0.18$ on average across all NIH Toolbox assessments). This includes the highest-performing 1D-CNN models (Table 3). We provide visualizations of the top 10 clusters for predicting individual cognitive performance using each input feature, along with the indices and anatomical labels of the clusters (Figure 3).

The top 10 clusters for different input features can be observed to belong to multiple different anatomical brain fiber tracts, including different types of tracts (superficial, association, cerebellar, striatal, and projection). However, commissural clusters (e.g.,

TABLE 3 | Pearson correlation coefficients (r) of the LASSO and 1D-CNN models. Values are the mean and standard deviation across the seven models trained to predict the seven NIH Toolbox cognitive assessments. The best result for each column is bolded. In each column, results that are not significantly different from the best result are in italics.

	LASSO regression AVG	CNN AVG
Microstructure		
FA	_0.157 ± 0.05_	_0.183 ± 0.1_
MD	_0.159 ± 0.04_	_0.174 ± 0.07_
Connectivity		
NoS	**0.18 ± 0.03**	_0.199 ± 0.09_
Shape		
Irregularity	_0.174 ± 0.03_	**0.2 ± 0.08**
Diameter	_0.148 ± 0.07_	_0.196 ± 0.09_
Total surface area	_0.171 ± 0.03_	_0.191 ± 0.09_
Branch volume	_0.152 ± 0.07_	_0.189 ± 0.09_
Volume	_0.174 ± 0.03_	_0.188 ± 0.09_
Surface area of end regions	_0.155 ± 0.05_	_0.181 ± 0.11_
Elongation	_0.175 ± 0.03_	_0.166 ± 0.07_
Radius of end regions	_0.149 ± 0.03_	0.164 ± 0.06
Curl	_0.153 ± 0.04_	0.145 ± 0.05
Span	_0.148 ± 0.04_	0.141 ± 0.08
Length	_0.145 ± 0.04_	0.125 ± 0.06
Trunk volume	_0.159 ± 0.03_	0.095 ± 0.02
Average	0.163 ± 0.01	0.169 ± 0.03

corpus callosum) are not identified in the top 10 most predictive clusters for any input feature in this experiment. In Figure 3, it can be observed that, for example, the top 10 most predictive fiber clusters for FA are all association fiber clusters, including nine superficial white matter fiber clusters (one is bilateral, superficial frontal cluster 604) and one left uncinate fiber cluster. As another example, the top 10 predictive fiber clusters for total surface area include 4 superficial white matter fiber clusters, 3 association clusters in the middle longitudinal fasciculus (bilateral cluster 691) and left cingulum, 2 intracerebellar parallel tract clusters (bilateral cluster 513), and a left striato-frontal projection cluster. The predictive clusters for the top-performing input feature, irregularity, span multiple fiber tract categories including 2 deep association fiber clusters (left cingulum), 4 superficial association fiber clusters, 2 cerebellar clusters (bilateral parallel tract cluster 516), 1 striatal cluster (left external capsule), and 1 projection cluster (left corona radiata frontal). Overall, this result suggests that fiber clusters with features predictive of cognitive ability are widespread throughout the brain and that different features capture different informative aspects of the white matter anatomy.
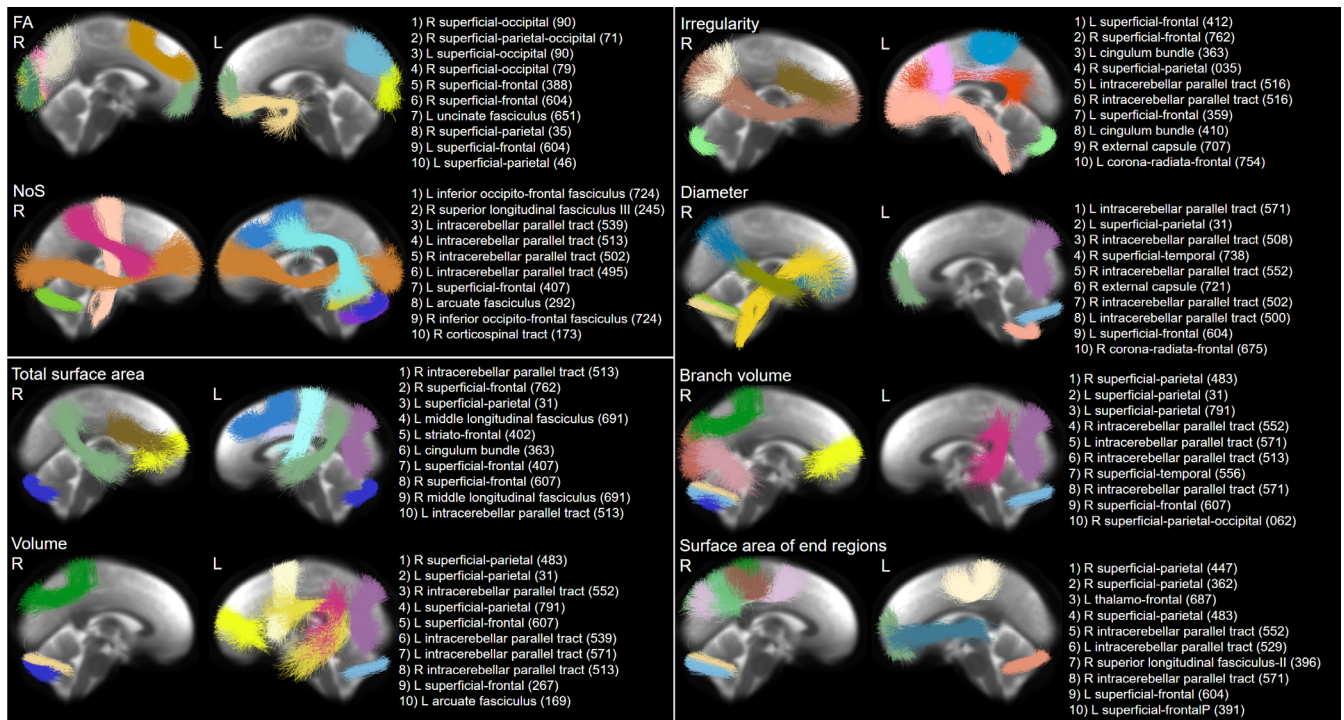
FIGURE 3 | Visualization of the top 10 most predictive fiber clusters for predicting individual cognitive performance using different input features. Fiber clusters are located within white matter fiber tracts, as listed to the right of the images (with fiber cluster atlas ID numbers provided in parentheses). Visualizations are performed using the fiber cluster atlas (Zhang, Wu, Norton et al. 2018).

4 | Discussion

This work has explored the potential of leveraging fiber cluster shape measures to predict subject-specific cognitive performance. We employed a data-driven strategy leveraging a large set of shape characteristics and assessments of individual cognitive performance. We applied two machine learning methods, 1D-CNN and LASSO regression. While both machine learning methods demonstrated that the proposed shape measures performed well in predicting individual cognitive performance, the 1D-CNN was more predictive on average. We introduced the widely used explainable approach, SHAP, to identify highly predictive fiber clusters. Our experimental results on a large-scale dataset of 1065 subjects demonstrated the effectiveness of shape measurements for individual cognitive performance prediction and identified white matter fiber clusters that highly contributed to the overall performance. Below, we discuss some detailed observations about our results in comparison with related studies in the literature.

Other groups have studied tractography shape measures over the past 20 years. Local shape measures, such as curvature and torsion, were proposed early in the field of tractography to describe local properties at points along individual streamlines (Batchelor et al. 2006; Corouge, Gouttard, and Gerig 2004). More recently, shape measures describing global properties of entire fiber tracts, such as the shape measures studied in this article, have received increasing interest. Multiple research groups have applied the DSIStudio shape computation software (Yeh 2020) to demonstrate that the shape of white matter fiber tracts varies in health and disease and across the lifespan (Linn et al. 2024; Schilling et al. 2022; Schilling, Archer

et al. 2023; Schilling, Chad et al. 2023; Yang et al. 2024; Yin et al. 2023).

Several recent studies have assessed potential relationships between white matter fiber tract shape and various measures of brain function, with a mix of positive and negative findings. These studies have generally focused on particular anatomical fiber tracts and their relationship with the brain functions the tracts are expected to subserve (for a review of fiber tracts and functions, see Forkel et al. 2022). For example, the shape (radius and irregularity of end regions) of the corticospinal tract was significantly associated with the risk of postoperative motor complications in low-grade glioma patients (Yang et al. 2024). In another example, the shape (volume and diameter) of the forceps minor of the corpus callosum was not associated with cognitive reappraisal, a mechanism for emotion regulation, in a healthy subjects dataset (Porcu et al. 2024). In a recent preprint, a study of the HCP-YA found that functional language lateralization was not significantly associated with shape metrics (including length, span, curl, elongation, diameter, volume, and surface area) of 11 language-related white matter tracts (Andrulyte et al. 2024). However, this study did not investigate assessments of language performance. Another recent study of the HCP-YA found a weak correlation of language performance assessments with a shape measure, the inferior end region coverage area of the frontal aslant tract (Linn et al. 2024). Overall, it can be noted that the above-related work provides only weak evidence that fiber tract shape features may relate to cognitive performance in healthy individuals. However, an initial conference publication by our group demonstrated the strong potential of shape features for machine-learning-based prediction of language performance (Lo, Chen et al. 2024). In contrast to these related

investigations, which studied a limited number of fiber tracts and cognitive assessments, the current study aimed to comprehensively evaluate multiple shape features, cognitive measures, and brain connections in the form of fiber clusters within fiber tracts. We sought to assess whether shape features predict cognitive abilities, which features are most predictive, and how their performance compares to traditional metrics like NoS and FA.

The results of the current study, which demonstrate that the investigated shape measures are predictive of cognitive performance, suggest that these shape measures are informative for studying the brain's white matter connections and their relationships to non-imaging phenotypes such as cognition. Furthermore, the results suggest that the study of shape features may be equally as informative as the study of traditional microstructure or connectivity features. Tables 1 and 2 provide clear evidence that the shape measures generally have similar predictive performance to traditional measures. While almost all shape measures perform reasonably well on average ($r > 0.1$), there is one clear exception where the Trunk Volume measure does not provide good performance when input to the 1D-CNN. The Trunk Volume was previously described as having low reliability in the test–retest sense (Yeh 2020). However, we note that the Trunk Volume measure performed well when using the LASSO models. Thus, our results suggest that all shape measures are informative not only about brain structure but also about the relationship between brain structure and brain function.

The application of explainable AI suggests that fiber clusters with features highly predictive of cognitive ability are widespread throughout the brain, including superficial association, deep association, cerebellar, striatal, and projection fiber clusters. The results further suggest that different features are able to capture different informative aspects of the white matter fiber cluster anatomy. As expected, some highly predictive fiber clusters (Figure 3) can be observed within fiber tracts expected to relate to aspects of cognition. For example, NoS and volume features of fiber clusters within the left arcuate fasciculus, classically critical for language function, are highly predictive of cognition. Similarly predictive are the total surface area and irregularity features of fiber clusters within the left cingulum bundle, a tract believed to be involved in multiple aspects of cognition (Forkel et al. 2022). Many features extracted from the fiber clusters within the superficial white matter, which provides cognitively critical cortico-cortical communication, are predictive of cognition (Wang et al. 2022). In Figure 3, it can also be observed that some clusters (e.g., superficial frontal cluster 607 and intracerebellar parallel tract cluster 571) are found to be highly predictive across several different features. This observation suggests that (as we already know) the shape features are not mathematically independent of each other. It also suggests that such consistently informative clusters may have the potential for future investigation in a more hypothesis-driven fashion.

The widespread nature of the predictive fiber clusters is in line with the understanding of cognition as arising from the interaction of multiple interconnected brain regions (Bressler and Menon 2010; Mesulam 1990; Uddin, Yeo, and Spreng 2019) and with a recent review demonstrating that fiber tracts are generally correlated with multiple cognitive domains (Forkel et al. 2022). While the biological mechanisms linking white matter shape

and cognitive performance are not fully understood, genetic and environmental factors influencing brain development may play a role (Luo et al. 2022). Differences in white matter shape could also be linked to structural changes in adjacent gray matter regions, including increased gyrification of neocortical regions, a factor associated with cognitive ability [Gregory et al. 2016; Lamballais et al. 2020; but see Mathias et al. 2020]. Overall, variability in shape features is likely to reflect numerous biological, environmental, genetic, and developmental factors that work in conjunction to influence the shape of an individual's brain and cognitive function.

We implemented the machine learning used in this article using two popular methods previously shown to be successful for the prediction of non-imaging phenotypes. Our methodological choices enabled us to perform the current computationally intensive, data-driven study. However, there are several limitations. The 1D-CNN may overlook potential synergies between shape, microstructure, and connectivity features. In future work, algorithms such as multi-view learning (Sun 2013; Wei et al. 2023; Xu, Tao, and Xu 2013) or ensembling learning (Gupta et al. 2017; Polikar 2012; Sagi and Rokach 2018; Zhang, Xue et al. 2022) could extract information from hidden correlations between features that remain unexplored when trained independently. Additionally, 1D-CNNs have proven effective in capturing local patterns in data but are inherently limited when modeling global features, especially in the context of high-dimensional data. Future work could benefit from exploring more sophisticated deep models, such as conditional diffusion-based models (Lo, Chen, Zhang et al. 2024; Yao et al. 2023) or transformer-based models (Chen, Zhang et al. 2024; Lo, Chen, Liu et al. 2024; Zhang, Xue et al. 2022). We implemented the explainable AI used in this article using a popular method, SHAP, that is widely applied across different machine learning approaches. In our study, we did not rely solely on the empirical SHAP values, as the values are not always informative in isolation. Instead, we ranked the SHAP values and averaged them to focus on predictive performance for overall cognition. Other approaches for organizing SHAP values, such as clustering (Clement et al. 2024; Durvasula et al. 2022), may provide additional insight. Future work could also investigate other explainable AI techniques, such as activation-based methods (Zhou et al. 2016), attention mechanisms (Devlin et al. 2018; Dong et al. 2021), gradient-based approaches (Selvaraju et al. 2017), or alternative perturbation methods (Ribeiro, Singh, and Guestrin 2016).

As one of the first studies to explore shape as a predictor of individual cognitive performance, this work has several key limitations and potential areas for future improvement. We investigated a large healthy YA dataset. Future work may investigate the effectiveness of shape measures in other datasets, such as those from different acquisitions or across the lifespan in health and disease. Future validation in such independent datasets (Casey et al. 2018; Cetin-Karayumak et al. 2024; Ge et al. 2023; Marek et al. 2011) can enable the assessment of the robustness and generalizability of the predictive models. Such large datasets can also enable investigation of the impact of potentially confounding factors such as education and socioeconomic status, which may influence performance on cognitive assessments and measures of the white matter. Furthermore, we relied on a single tractography method and a single atlas-based definition of major fiber tracts (Zhang, Savadjiev

et al. [2018](); Zhang, Wu, Ning et al. [2018](); Zhang, Wu, Norton et al. [2018]()). An in-depth comparison of tractography methods and white matter tract definitions (He et al. [2023](); Schilling et al. [2021]()) remains an interesting future work to obtain a more robust understanding of the effectiveness of shape measures in different methodological scenarios. We also relied on a single method for interpretation, SHAP. In future work, combining or comparing additional model-agnostic interpretation approaches may offer alternative insights (Apley and Zhu [2020](); Ribeiro, Singh, and Guestrin [2016](); Ribeiro, Singh, and Guestrin [2018]()). In addition, it remains to be investigated if shape may predict other non-imaging phenotypes beyond NIH Toolbox assessments of cognition. Future work could explore other structure–function relationships, using additional non-imaging phenotypes and potentially more sophisticated machine learning and explainable AI approaches.

5 | Conclusion

In this work, we investigated fiber cluster shape features, a novel approach for cognitive performance prediction. Using explainable deep learning, we demonstrated that the shape of the brain's connections is predictive of individual cognitive performance. We showed that shape can achieve better or equivalent prediction performance on cognitive prediction in comparison to conventional measurements of microstructure and connectivity. Our results suggest that all shape measures under study are relevant for the study of the brain's white matter connections and their relationships to non-imaging phenotypes such as cognition. Overall, this study demonstrates the potential of utilizing geometric shape descriptors of tractography fiber clusters to enhance the study of the brain's white matter and its relationship to various cognitive functions.

Ethics Statement

The creation of the WU-Minn HCP dataset was approved by the institutional review board of Washington University in St. Louis (IRB #201204036).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The Human Connectome Project minimally preprocessed young adult dataset imaging data and associated NIH Toolbox measures are publicly available at https://db.humanconnectome.org/. The ORG tractography atlas is publicly available at http://dmri.slicer.org/atlases/, and code to apply the atlas is publicly available at https://github.com/SlicerDMRI/whitematteranalysis. All code developed for our experiments will be publicly available at https://github.com/SlicerDMRI/TractShapeCognition.

Endnotes

[1] https://github.com/pnlbwh/ukftractography.

[2] https://github.com/SlicerDMRI/whitematteranalysis.

[3] https://dsi-studio.labsolver.org/.

References

Andrulyte, I., C. De Bezenac, F. Branzi, S. J. Forkel, P. N. Taylor, and S. S. Keller. 2024. "The Relationship Between White Matter Architecture and Language Lateralisation in the Healthy Brain. bioRxiv." https://doi.org/10.1101/2024.01.12.572159v2.

Apley, D. W., and J. Zhu. 2020. "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 82: 1059–1086.

Azevedo, T., L. Passamonti, P. Lió, and N. Toschi. 2019. "A Machine Learning Tool for Interpreting Differences in Cognition Using Brain Features." In *Artificial Intelligence Applications and Innovations*, 475–486. Springer International Publishing.

Ballester, P. L., J. S. Suh, N. C. W. Ho, et al. 2023. "Gray Matter Volume Drives the Brain Age Gap in Schizophrenia: A SHAP Study." *Schizophrenia* 9, no. 3: 3.

Basser, P. J., S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. 2000. "In Vivo Fiber Tractography Using DT-MRI Data." *Magnetic Resonance in Medicine* 44: 625–632.

Batchelor, P. G., F. Calamante, J.-D. Tournier, D. Atkinson, D. L. G. Hill, and A. Connelly. 2006. "Quantification of the Shape of Fiber Tracts." *Magnetic Resonance in Medicine* 55: 894–903.

Bauer, P. J., S. S. Dikmen, R. K. Heaton, D. Mungas, J. Slotkin, and J. L. Beaumont. 2013. "III. NIH Toolbox Cognition Battery (CB): Measuring Episodic Memory." *Monographs of the Society for Research in Child Development* 78: 34–48.

Beaulieu, C. 2009. "The Biological Basis of Diffusion Anisotropy." In *Diffusion MRI*, 105–126. Elsevier.

Bressler, S. L., and V. Menon. 2010. "Large-Scale Brain Networks in Cognition: Emerging Methods and Principles." *Trends in Cognitive Sciences* 14: 277–290.

Carlozzi, N. E., J. L. Beaumont, D. S. Tulsky, and R. C. Gershon. 2015. "The NIH Toolbox Pattern Comparison Processing Speed Test: Normative Data." *Archives of Clinical Neuropsychology* 30: 359–368.

Casey, B. J., T. Cannonier, M. I. Conley, et al. 2018. "The Adolescent Brain Cognitive Development (ABCD) Study: Imaging Acquisition Across 21 Sites." *Developmental Cognitive Neuroscience* 32: 43–54.

Cetin-Karayumak, S., F. Zhang, R. Zurrin, et al. 2024. "Harmonized Diffusion MRI Data and White Matter Measures From the Adolescent Brain Cognitive Development Study." *Scientific Data* 11: 249.

Chamberland, M., S. Genc, C. M. W. Tax, et al. 2021. "Detecting Microstructural Deviations in Individuals With Deep Diffusion MRI Tractometry." *Nature Computational Science* 1: 598–606.

Chen, M., H. Li, J. Wang, et al. 2020. "Early Prediction of Cognitive Deficit in Very Preterm Infants Using Brain Structural Connectome With Transfer Learning Enhanced Deep Convolutional Neural Networks." *Frontiers in Neuroscience* 14: 858.

Chen, Y., L. R. Zekelman, C. Zhang, et al. 2024. "TractGeoNet: A Geometric Deep Learning Framework for Pointwise Analysis of Tract

Microstructure to Predict Language Assessment Performance." *Medical Image Analysis* 94: 103120.

Chen, Y., F. Zhang, M. Wang, et al. 2024. "TractGraphFormer: Anatomically informed hybrid Graph CNN-Transformer network for classification from diffusion MRI tractography. arXiv [cs.CV]. arXiv." http://arxiv.org/abs/2407.08883.

Chen, Y., F. Zhang, L. R. Zekelman, et al. 2023. "Tractgraphcnn: Anatomically Informed Graph CNN for Classification Using Diffusion MRI Tractography." In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI),* 1–5. IEEE.

Chen, Y., F. Zhang, C. Zhang, et al. 2022. "White Matter Tracts Are Point Clouds: Neuropsychological Score Prediction and Critical Region Localization via Geometric Deep Learning." In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*, 174–184. Springer Nature Switzerland.

Clement, T., H. T. T. Nguyen, N. Kemmerzell, M. Abdelaal, and D. Stjelja. 2024. "Beyond Explaining: XAI-Based Adaptive Learning With SHAP Clustering for Energy Consumption Prediction. arXiv [cs.LG]. arXiv." http://arxiv.org/abs/2402.04982.

Corouge, I., S. Gouttard, and G. Gerig. 2004. "Towards a Shape Model of White Matter Fiber Bundles Using Diffusion Tensor MRI." In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, 344–347. IEEE.

Cui, Z., and G. Gong. 2018. "The Effect of Machine Learning Regression Algorithms and Sample Size on Individualized Behavioral Prediction With Functional Connectivity Features." *NeuroImage* 178: 622–637.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. "BERT: Pre-training of deep bidirectional Transformers for language understanding. arXiv [cs.CL]. arXiv." http://arxiv.org/abs/1810.04805.

Dong, Y., C. Liu, Y. Lo, Y. Xu, K. Wang, and K. Zhang. 2021. "Attention Backpropagation: A Backtracking Approach to Visualize the Discriminative Image Regions." In *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence*. New York, NY: ACM. https://dl.acm.org/doi/abs/10.1145/3448218.3448227.

Durvasula, N., V. d'Hauteville, K. Hines, and J. P. Dickerson. 2022. "Characterizing Anomalies With Explainable Classifiers. In NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications." https://openreview.net/pdf?id=P_ImdNqo7S.

Feng, G., Y. Wang, W. Huang, et al. 2022. "Methodological Evaluation of Individual Cognitive Prediction Based on the Brain White Matter Structural Connectome." *Human Brain Mapping* 43: 3775–3791.

Forkel, S. J., P. Friedrich, M. Thiebaut de Schotten, and H. Howells. 2022. "White Matter Variability, Cognition, and Disorders: A Systematic Review." *Brain Structure & Function* 227: 529–544.

Gabusi, I., M. Battocchio, S. Bosticardo, S. Schiavi, and A. Daducci. 2024. "Blurred Streamlines: A Novel Representation to Reduce Redundancy in Tractography." *Medical Image Analysis* 93: 103101.

Ge, J., G. Yang, M. Han, et al. 2023. "Increasing Diversity in Connectomics With the Chinese Human Connectome Project." *Nature Neuroscience* 26: 163–172.

Gershon, R. C., K. F. Cook, D. Mungas, et al. 2014. "Language Measures of the NIH Toolbox Cognition Battery." *Journal of the International Neuropsychological Society* 20: 642–651.

Gong, W., C. F. Beckmann, and S. M. Smith. 2021. "Phenotype Discovery From Population Brain Imaging." *Medical Image Analysis* 71: 102050.

Gregory, M. D., J. S. Kippenhan, D. Dickinson, et al. 2016. "Regional Variations in Brain Gyrification Are Associated With General Cognitive Ability in Humans." *Current Biology* 26: 1301–1305.

Gupta, V., S. I. Thomopoulos, F. M. Rashid, and P. M. Thompson. 2017. "FiberNET: An Ensemble Deep Learning Framework for Clustering White Matter Fibers." In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*, Lecture Notes in Computer Science, 548–555. Cham: Springer International Publishing.

He, H., F. Zhang, S. Pieper, et al. 2022. "Model and Predict Age and Sex in Healthy Subjects Using Brain White Matter Features: A Deep Learning Approach. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). ieeexplore.ieee.org. pp 1-5."

He, J., F. Zhang, Y. Pan, et al. 2023. "Reconstructing the Somatotopic Organization of the Corticospinal Tract Remains a Challenge for Modern Tractography Methods." *Human Brain Mapping* 44: 6055–6073.

Hodes, R. J., T. R. Insel, S. C. Landis, and NIH Blueprint for Neuroscience Research. 2013. "The NIH Toolbox: Setting a Standard for Biomedical Research." *Neurology* 80: S1.

Jeong, J.-W., M.-H. Lee, N. O'Hara, C. Juhász, and E. Asano. 2021. "Prediction of Baseline Expressive and Receptive Language Function in Children With Focal Epilepsy Using Diffusion Tractography-Based Deep Learning Network." *Epilepsy & Behavior* 117: 107909.

Jones, D. K., T. R. Knösche, and R. Turner. 2013. "White Matter Integrity, Fiber Count, and Other Fallacies: The do's and don'ts of Diffusion MRI." *NeuroImage* 73: 239–254.

Keller, C. J., S. Bickel, L. Entz, et al. 2011. "Intrinsic Functional Architecture Predicts Electrically Evoked Responses in the Human Brain." *Proceedings of the National Academy of Sciences of the United States of America* 108: 10308–10313.

Kim, M., J. Bao, K. Liu, et al. 2021. "A Structural Enriched Functional Network: An Application to Predict Brain Cognitive Performance." *Medical Image Analysis* 71: 102026.

Lamballais, S., E. J. Vinke, M. W. Vernooij, M. A. Ikram, and R. L. Muetzel. 2020. "Cortical Gyrification in Relation to Age and Cognition in Older Adults." *NeuroImage* 212: 116637.

Lebel, C., M. Gee, R. Camicioli, M. Wieler, W. Martin, and C. Beaulieu. 2012. "Diffusion Tensor Imaging of White Matter Tract Evolution Over the Lifespan." *NeuroImage* 60: 340–352.

Linn, W.-J., J. Barrios-Martinez, D. Fernandes-Cabral, et al. 2024. "Probabilistic Coverage of the Frontal Aslant Tract in Young Adults: Insights Into Individual Variability, Lateralization, and Language Functions." *Human Brain Mapping* 45: e26630.

Liu, R., M. Li, and D. B. Dunson. 2023. "PPA: Principal Parcellation Analysis for Brain Connectomes and Multiple Traits." *NeuroImage* 276: 120214.

Liu, W., Y. Chen, C. Ye, et al. 2023. "Fiber Tract Shape Measures Inform Prediction of Non-Imaging Phenotypes. arXiv [cs.CV]. arXiv." http://arxiv.org/abs/2303.09124.

Lo, Y., Y. Chen, D. Liu, et al. 2024. "Cross-Domain Fiber Cluster Shape Analysis for Language Performance Cognitive Score Prediction. ArXiv." https://www.ncbi.nlm.nih.gov/pubmed/38584619.

Lo, Y., Y. Chen, F. Zhang, et al. 2024. "White Matter Geometry-Guided Score-Based Diffusion Model for Tissue Microstructure Imputation in Tractography Imaging. arXiv [cs.CV]. arXiv." http://arxiv.org/abs/2407.19460.

Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777. Curran Associates Inc.

Luo, Z., N. Adluru, D. C. Dean 3rd, A. L. Alexander, and H. H. Goldsmith. 2022. "Genetic and Environmental Influences of Variation in Diffusion MRI Measures of White Matter Microstructure." *Brain Structure & Function* 227: 131–144.

Marek, K., D. Jennings, S. Lasch, et al. 2011. "The Parkinson Progression Marker Initiative (PPMI)." *Progress in Neurobiology* 95: 629–635.

Mathias, S. R., E. E. M. Knowles, J. Mollon, et al. 2020. "Minimal Relationship Between Local Gyrification and General Cognitive Ability in Humans." *Cerebral Cortex* 30: 3439–3450.

Mesulam, M. M. 1990. "Large-Scale Neurocognitive Networks and Distributed Processing for Attention, Language, and Memory." *Annals of Neurology* 28: 597–613.

Nie, X., J. Ruan, M. C. G. Otaduy, L. T. Grinberg, J. Ringman, and Y. Shi. 2024. "Surface-Based Probabilistic Fiber Tracking in Superficial White Matter." *IEEE Transactions on Medical Imaging* 43: 1113–1124.

O'Donnell, L. J., and C.-F. Westin. 2007. "Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas." *IEEE Transactions on Medical Imaging* 26: 1562–1575.

Paszke, A., S. Gross, F. Massa, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, vol. 721, 8026–8037. Red Hook, NY: Curran Associates.

Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.

Polikar, R. 2012. "Ensemble Learning." In *Ensemble Machine Learning*, 1–34. New York, NY: Springer New York.

Porcu, M., L. Cocco, R. Cau, et al. 2024. "Correlation of Cognitive Reappraisal and the Microstructural Properties of the Forceps Minor: A Deductive Exploratory Diffusion Tensor Imaging Study." *Brain Topography* 37: 63–74.

Qi, S., S. Meesters, K. Nicolay, B. M. T. H. Romeny, and P. Ossenblok. 2015. "The Influence of Construction Methodology on Structural Brain Network Measures: A Review." *Journal of Neuroscience Methods* 253: 170–182.

Rasero, J., A. I. Sentis, F.-C. Yeh, and T. Verstynen. 2021. "Integrating Across Neuroimaging Modalities Boosts Prediction Accuracy of Cognitive Ability." *PLoS Computational Biology* 17: e1008347.

Reddy, C. P., and Y. Rathi. 2016. "Joint Multi-Fiber NODDI Parameter Estimation and Tractography Using the Unscented Information Filter." *Frontiers in Neuroscience* 10: 166.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "Why Should I Trust You?: Explaining the Predictions of any Classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 16–1135. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2018. "Anchors: High-Precision Model-Agnostic Explanations. Proceedings of the AAAI Conference on Artificial Intelligence 32." https://doi.org/10.1609/aaai.v32i1.11491.

Sagi, O., and L. Rokach. 2018. "Ensemble Learning: A Survey." *WIREs Data Mining and Knowledge Discovery* 8: e1249.

Scheda, R., and S. Diciotti. 2022. "Explanations of Machine Learning Models in Repeated Nested Cross-Validation: An Application in Age Prediction Using Brain Complexity Features." *NATO Advanced Science Institutes Series E: Applied Sciences* 12: 6681.

Schilling, K. G., D. Archer, F.-C. Yeh, et al. 2022. "Aging and White Matter Microstructure and Macrostructure: A Longitudinal Multi-Site Diffusion MRI Study of 1218 Participants." *Brain Structure & Function* 227: 2111–2125.

Schilling, K. G., D. Archer, F.-C. Yeh, et al. 2023. "Short Superficial White Matter and Aging: A Longitudinal Multi-Site Study of 1293 Subjects and 2711 Sessions." *Aging Brain* 3: 100067.

Schilling, K. G., J. A. Chad, M. Chamberland, et al. 2023. "White Matter Tract Microstructure, Macrostructure, and Associated Cortical Gray Matter Morphology Across the Lifespan." *Imaging Neuroscience* 1: 1–24.

Schilling, K. G., F. Rheault, L. Petit, et al. 2021. "Tractography Dissection Variability: What Happens When 42 Groups Dissect 14 White Matter Bundles on the Same Dataset?" *NeuroImage* 243: 118502.

Sedgwick, P. 2012. "Pearson's Correlation Coefficient." *BMJ* 345: e4483. https://doi.org/10.1136/bmj.e4483.

Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. 2017. "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization." In *2017 IEEE International Conference on Computer Vision (ICCV)*, 618–626. IEEE.

Shen, T., C. Li, B. Wang, et al. 2015. "Increased Cognition Connectivity Network in Major Depression Disorder: A FMRI Study." *Psychiatry Investigation* 12: 227–234.

Sun, J., Z. Tu, D. Meng, Y. Gong, M. Zhang, and J. Xu. 2022. "Interpretation for Individual Brain Age Prediction Based on Gray Matter Volume." *Brain Sciences* 12, no. 11: 1517. https://doi.org/10.3390/brainsci12111517.

Sun, S. 2013. "A Survey of Multi-View Machine Learning." *Neural Computing and Applications* 23: 2031–2038.

Tian, Y., and A. Zalesky. 2021. "Machine Learning Prediction of Cognition From Functional Connectivity: Are Feature Weights Reliable?" *NeuroImage* 245: 118648.

Tibshirani, R. 2018. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 58: 267–288.

Tobyne, S. M., D. C. Somers, J. A. Brissenden, S. W. Michalka, A. L. Noyce, and D. E. Osher. 2018. "Prediction of Individualized Task Activation in Sensory Modality-Selective Frontal Cortex With 'Connectome Fingerprinting'." *NeuroImage* 183: 173–185.

Tulsky, D. S., N. Carlozzi, N. D. Chiaravalloti, et al. 2014. "NIH Toolbox Cognition Battery (NIHTB-CB): List Sorting Test to Measure Working Memory." *Journal of the International Neuropsychological Society* 20: 599–610.

Uddin, L. Q., B. T. T. Yeo, and R. N. Spreng. 2019. "Towards a Universal Taxonomy of Macro-Scale Functional Human Brain Networks." *Brain Topography* 32: 926–942.

Van Essen, D. C., S. M. Smith, D. M. Barch, et al. 2013. "The WU-Minn Human Connectome Project: An Overview." *NeuroImage* 80: 62–79.

Van Essen, D. C., K. Ugurbil, E. Auerbach, et al. 2012. "The Human Connectome Project: A Data Acquisition Perspective." *NeuroImage* 62: 2222–2231.

Wang, S., F. Zhang, P. Huang, et al. 2022. "Superficial White Matter Microstructure Affects Processing Speed in Cerebral Small Vessel Disease." *Human Brain Mapping* 43: 5310–5325.

Wei, Y., Y. Chen, T. Xue, et al. 2023. "A Deep Network for Explainable Prediction of Non-imaging Phenotypes Using Anatomical Multi-View Data." In *Computational Diffusion MRI*, 165–176. Springer Nature Switzerland.

Weintraub, S., S. S. Dikmen, R. K. Heaton, et al. 2013. "Cognition Assessment Using the NIH Toolbox." *Neurology* 80: S54–S64.

Wen, Q., S. M. Mustafi, J. Li, et al. 2019. "White Matter Alterations in Early-Stage Alzheimer's Disease: A Tract-Specific Study." *Alzheimers Dement* 11: 576–587.

Wu, J., J. Li, S. B. Eickhoff, D. Scheinost, and S. Genon. 2023. "The Challenges and Prospects of Brain-Based Prediction of Behaviour." *Nature Human Behaviour* 7: 1255–1264.

Xiao, Y., Y. Lin, J. Ma, et al. 2021. "Predicting Visual Working Memory With Multimodal Magnetic Resonance Imaging." *Human Brain Mapping* 42: 1446–1462.

Xu, C., D. Tao, and C. Xu. 2013. "A Survey on Multi-View Learning. arXiv [cs.LG]. arXiv." http://arxiv.org/abs/1304.5634.

Xue, T., F. Zhang, L. R. Zekelman, et al. 2024. "TractoSCR: A Novel Supervised Contrastive Regression Framework for Prediction of Neurocognitive Measures Using Multi-Site Harmonized Diffusion MRI Tractography." *Frontiers in Neuroscience* 18: 1411797. https://doi.org/10.3389/fnins.2024.1411797/full.

Yang, Z.-C., F.-C. Yeh, B.-W. Xue, et al. 2024. "Assessing Postoperative Motor Risk in Insular Low-Grade Gliomas Patients: The Potential Role of Presurgery MRI Corticospinal Tract Shape Measures." *Journal of Magnetic Resonance Imaging* 60, no. 5: 1892–1901. https://doi.org/10.1002/jmri.29244.

Yao, W., Y. Shen, F. Nicolls, and S.-Q. Wang. 2023. "Conditional Diffusion Model-Based Data Augmentation for Alzheimer's Prediction." In *Communications in Computer and Information Science*, 33–46. Singapore: Springer Nature Singapore.

Yeh, F.-C. 2020. "Shape Analysis of the Human Association Pathways." *NeuroImage* 223: 117329.

Yin, H., F. Zong, X. Deng, et al. 2023. "The Language-Related Cerebro-Cerebellar Pathway in Humans: A Diffusion Imaging-Based Tractographic Study." *Quantitative Imaging in Medicine and Surgery* 13: 1399–1416.

Zekelman, L. R., F. Zhang, N. Makris, et al. 2022. "White Matter Association Tracts Underlying Language and Theory of Mind: An Investigation of 809 Brains From the Human Connectome Project." *NeuroImage* 246: 118739.

Zelazo, P. D., J. E. Anderson, J. Richler, K. Wallner-Allen, J. L. Beaumont, and S. Weintraub. 2013. "II. NIH Toolbox Cognition Battery (CB): Measuring Executive Function and Attention." *Monographs of the Society for Research in Child Development* 78: 16–33.

Zhang, F., A. Daducci, Y. He, et al. 2022. "Quantitative Mapping of the Brain's Structural Connectivity Using Diffusion MRI Tractography: A Review." *NeuroImage* 249: 118870.

Zhang, F., P. Savadjiev, W. Cai, et al. 2018. "Whole Brain White Matter Connectivity Analysis Using Machine Learning: An Application to Autism." *NeuroImage* 172: 826–837.

Zhang, F., W. Wu, L. Ning, et al. 2018. "Suprathreshold Fiber Cluster Statistics: Leveraging White Matter Geometry to Enhance Tractography Statistical Analysis." *NeuroImage* 171: 341–354.

Zhang, F., Y. Wu, I. Norton, Y. Rathi, A. J. Golby, and L. J. O'Donnell. 2019. "Test-Retest Reproducibility of White Matter Parcellation Using Diffusion MRI Tractography Fiber Clustering." *Human Brain Mapping* 40: 3041–3057.

Zhang, F., Y. Wu, I. Norton, et al. 2018. "An Anatomically Curated Fiber Clustering White Matter Atlas for Consistent White Matter Tract Parcellation Across the Lifespan." *NeuroImage* 179: 429–447.

Zhang, F., T. Xue, W. Cai, Y. Rathi, C.-F. Westin, and L. J. O'Donnell. 2022. "TractoFormer: A Novel Fiber-Level Whole Brain Tractography Analysis Framework Using Spectral Embedding and Vision Transformers." In *Lecture Notes in Computer Science*, 196–206. Cham: Springer Nature Switzerland.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. 2016. "Learning Deep Features for Discriminative Localization." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2016.319.

Zou, H., and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 67: 301–320.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.