2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Altered Connection and Diagnosis Utility of White Matter in Alzheimer's Disease: A Multi-site Automated Fiber Quantification Study

Yida Qu, Pan Wang, Bing Liu, Xiaopeng Kang, Pindong Chen, Kai Du, Yong Liu, *Member, IEEE*

***Abstract*— Alzheimer's disease (AD) is a typical neurodegenerative disease that is associated with cognitive decline, memory loss, and functional disconnection. Diffusion tensor imaging (DTI) has been widely used to investigate the integrity and degeneration of white matter in AD. In this study, with one of the world's largest DTI biobanks (865 individuals), we aim to explore the diagnosis utility and stability of tract-based features (extracted by automated fiber quantification (AFQ) pipeline) in AD. First, we studied the clinical association of tract-based features by detecting AD-associated alterations of diffusion properties along fiber bundles. Then, a binary classification experiment between AD and normal controls was performed using tract-based diffusion properties as features and support vector machine (SVM) as a classifier with an independent site cross-validation strategy. The average accuracy of 77.90% (the highest was 88.89%) showed that white matter properties as biomarkers had a relatively stable role in the clinical diagnosis of AD.**

***Clinical Relevance*— White matter characteristics are valid and robust biomarkers of AD, which have high accuracy and generalizability in the AD diagnosis in a large multi-site dataset.**

I. Introduction

Alzheimer's disease (AD) is a chronic, progressive neurodegenerative disease associated with cognitive decline, which brings a huge burden to both family and society [1, 2]. Numerous studies have concluded that effective disease control benefits from early diagnosis and intervention, which gives a research priority to the biomarker searching for early diagnosis of AD [3, 4].

Several studies have shown that white matter (WM) impairment is a part of the pathological cascade in AD [5, 6]. Diffusion tensor imaging (DTI) can provide quantitative measurement of WM fiber trend and damage degree, providing valuable information for WM lesions. Among DTI analysis methods, automated fiber quantification (AFQ) has shown certain advantages in processing large-scale data efficiently and detecting detailed information based on the points along fiber bundles automatically [7, 8]. Thus AFQ has been used to study the tracts alterations in several neurogenic diseases including AD [9-12].

Dou and colleagues had introduced the AFQ method to study the WM alterations in AD and explored the potential possibility of using WM indices to classify AD from normal controls (NC) [12]. But this study was based on a small data set thus lacked persuasiveness. For searching imaging biomarkers with potentiality in the clinical application of AD, it is of great significance to use cross-validation based on large-scale multi-site databases to improve the generalizability of the algorithmic models. Hereby, the current study aims to use a large-scale multisite dataset to investigate the clinical validity and stability of the tract abnormal patterns using a consistent analysis pipeline, as well as to study the diagnosis ability and generalizability of WM features in classifying AD from NC among different sites.

For this purpose, in the current study, based on a multi-site dataset with a total of 865 subjects (321 ADs, 265 mild cognitive impairment (MCI) subjects, 279 NCs), firstly, we tested the clinical association of the tract-based features extracted via AFQ through a meta-analysis, which was to detect pointwise differences between AD and NC groups of diffusion parameters (fractional anisotropy (FA), radial diffusivity (RD), mean diffusivity (MD), and axial diffusivity (AxD)), followed by correlation analysis between these parameters and the score of cognitive status (mini-mental state examination (MMSE)) in the MCI plus AD groups. Secondly, ADs and NCs were binary classified using tract-based properties as features and support vector machine (SVM) as a classifier with independent site cross-validation to estimate the diagnosis utility and generalizability of tract-based features in AD.

II. Methods

A. Participants and Image Acquisition Protocol

Statement: This study was approved by the Medical Research Ethics Committee and Institutional Review Board.

The present study contained a total of 865 subjects (321 ADs, 265 MCIs, 279 NCs) from 7 different MRI machines in 4 hospitals in China. The acquisition parameters of DTI images were listed in Table I [13].

* This work was partially supported by the National Key Research and Development Program of China (No. 2017YFB1002502), the National Natural Science Foundation of China (Grant No. 81871438, 81901101).

Y. Qu, P. Chen, X. Kang, K. Du are with Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. They are also with University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: quyida2019@ia.ac.cn, chenpindong2019@ia.ac.cn, kangxiaopeng2018@ia.ac.cn, dukai2019@ia.ac.cn).

P. Wang is with Department of Neurology, Tianjin Huanhu Hospital, Tianjin University, Tianjin 300350, China (e-mail: wpaofeier@163.com).

B. Liu is with State Key Lab of Cognition Neuroscience & Learning, Beijing Normal University, Beijing 100875, China (e-mail: bing.liu@bnu.edu.cn).

Y. Liu is the faculty of the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China (corresponding to Y. Liu: e-mail: yongliu@bupt.edu.cn).

TABLE I. DTI ACQUISITION PARAMETERS

Study	S01	S02	S03	S04	S05	S06	S07
MRI strength	3.0T	3.0T	3.0T	3.0T	3.0T	3.0T	3.0T
Brand	Siemens Skyra	GE Signa HDx	Siemens Trio Tim	Siemens Verio	Siemens Trio Tim	Siemens Trio Tim	Siemens Skyra
Directions	64	25	64	64	30	12	64
B-values(s/mm^2)	1000	1000	1000	1000	1000	1000	1000
Repetition Time (ms)	7000	8000	6000	6400	11000	6000	7000
Echo time (ms)	91	88	85	98	98	87	91
Matrix	128*128	256*256	128*128	128*128	128*116	256*256	128*128
Field of view	256*256	220*220	256*256	256*256	256*232	256*256	256*256
Slice number /thickness (mm)	50/3	29/4	45/3	45/3	60/2	45/3	50/3

B. Image Preprocessing and WM Features Extraction

Images of all datasets were processed by the following procedures. Firstly, DICOM files were transformed to the Nifti images. Then, the DWI images were processed for routine dtiInit preprocessing pipeline in the VISTASOFT package (version 1.0, https://github.com/vistalab/vistasoft) containing eddy current correction, head motion correction, skull stripping, and tensor model fitting. Secondly, the AFQ pipeline (version 1.2, https://github.com/yeatmanlab/AFQ) [7] was used to quantify diffusion parameters (FA, RD, MD, AxD) and morphometric characteristics (fiber core linearity values, curvature, torsion, volume) at 100 points along the 18 major white matter tracts, the names of these tracts including bilateral inferior fronto-occipital fasciculus (IFOF), corticospinal, inferior longitudinal fasciculus (ILF), arcuate, superior longitudinal fasciculus (SLF), cingulum cingulate, thalamic radiation, uncinate fasciculus, minor and major callosum forceps.

Due to complex factors such as data quality, there is no guarantee of the recognition of all 18 fiber bundles for each subject using AFQ. Thus, during the quality control step, subjects with no more than 14 fiber bundles identified were excluded from the subsequent analysis. Consequently, 825 subjects (276 NCs, 255 MCIs, 294 ADs) were included for further analysis (Table II). In addition, curvature and torsion were excluded from subsequent analysis due to excessive missing values.

TABLE II. DEMOGRAPHICS AND PSYCHOLOGICAL STATISTICS OF THE INCLUDED SUBJECTS

Site	NC	MCI	AD	ALL	Age, mean±SD				Gender, male/female			MMSE, mean±SD			
					NC	*MCI*	*AD*		*NC*	*MCI*	*AD*	*NC*	*MCI*	*AD*	
S01	43	35	44	122	68.4 ±6.7	69.8 ±8.8	70.1 ±8.8		20/23	14/21	21/23	28.6 ±1.4	26.5 ±2.6	17.6 ±6.3	***
S02	26	22	30	78	69.3 ±6.6	73.5 ±8.5	72.1 ±8.3		15/11	12/10	9/21	28.9 ±1.0	26.9 ±1.7	19.9 ±4.2	***
S03	24	35	40	99	65.5 ±6.2	65.5 ±8.4	66.7 ±8.2		9/15	11/24	20/20	28.8 ±1.2	25.5 ±3.2	15.5 ±5.8	***
S04	42	16	65	123	65.5 ±6.8	66.1 ±7.4	67.8 ±7.0		12/30	8/8	29/36	28.5 ±1.7	24.8 ±1.5	18.9 ±3.3	***
S05	79	96	47	222	63.7 ±7.8	67.6 ±10.2	69.2 ±9.2	**	30/49	49/47	17/30	28.2 ±2.3	24.2 ±3.6	17.1 ±6.4	***
S06	19	14	26	59	65.1 ±8.4	70.9 ±9.6	63.9 ±7.6	*	7/12	10/4	11/15	28.3 ±1.6	21.5 ±5.5	10.4 ±6.7	***
S07	43	37	42	122	68.3 ±8.0	69.8 ±6.9	70.4 ±8.7		20/23	13/24	16/26	28.8 ±1.2	26.3 ±2.6	15.8 ±5.7	***

Note: Gender comparisons between groups were tested through Chi-squared tests. One-way ANOVA was performed for age and MMSE comparisons. ***p<0.001, **P<0.01, *P<0.05.

C. Group-level Pointwise Meta-analysis

The clinical association of tract-based features was investigated by detecting reliable differences of these features between the AD and NC groups and the relationship with cognitive states. First, we performed a pointwise analysis of the diffusion metrics along 100 points of each fiber tract at a group level (AD and NC). After controlling age and gender effects, a random-effects meta-analysis (R, meta-package, version 4.9-7) was used to combine effect sizes (Cohen's d) of 7 cohorts for each diffusion property and each point along the right and left tracts separately. Points with P<0.01 (FDR corrected) and fiber tracts with at least 10 continuous and uninterrupted significant points were considered statistically significant between AD and NC groups.

Then, we calculated the Pearson's correlation coefficients to investigate the relationship between cognitive ability scores (MMSE) and diffusion parameters that were averaged over the intervals with significant differences (results of meta-analysis) at each major tract in the MCI plus AD groups (P<0.005, uncorrected).

D. Classification

After verifying the clinical validity of tract-based features through meta-analysis and correlation analysis, we performed classification experiments to evaluate the discriminant and generalization performance of tract-based features in AD diagnosis. We used tract-based properties extracted via AFQ as features and SVM with radial basis function (RBF) kernel (python 3.7, Scikit-learn package 0.21.2) as a classifier to achieve a binary classification task (276 NCs and 294 ADs).

Specifically, we first reduced the feature dimension from 10802 (18fibers×100points×6metrics (FA, MD, RD, AxD, fiber core linearity values and volume) + age and gender = 10802) to 1082 by applying a 10-point-average smooth strategy to tract-based features. Then, for the outer loop, we adopted a leave-one-site-out cross-validation strategy which means that one site data was selected as the testing set for each loop, data from the rest of the sites were the training set. Within the inner-loop, we used the training set to search optimal model parameters for SVM (penalty parameter C of the error term and RBF kernel coefficient γ) using grid search method with 3-fold cross-validation (the default fold number in Scikit-learn package 0.21.2) (Fig. 2A) [14]. Lastly, Accuracy (ACC), sensitivity (SEN), and specificity (SPE) as well as the area under the receiver operating characteristic curve (AUC) were used to evaluate model performance. Besides, Pearson's correlation between the classifier output (decision function score) and MMSE was calculated to find the clinical relevance of classification. In particular, the absolute value of the decision function score represents the credibility of prediction and the greater absolute value means higher credibility. The sign of the decision function score represents the predicted class (plus sign represents AD and minus sign represents NC).

In addition, considering the heterogeneity of DTI data from different sites, we harmonized the tract-based features with the Combat method (combat tool, sva package, version 3.34.0 in R) [15] and then performed the classification task in the same way as a supplementary experiment. We also used XGBoost and linear kernel SVM as classifiers to repeat the experiment to study the feature importance scores.

III. Results

A. *Meta-analysis of Pointwise Differences along Tracts Showed the Clinical Validity of the Tract-based Features*

Meta-analysis results showed that differences between AD and NC groups widely existed in tract-based features, and the diffusion properties of the discrepant loci were significantly correlated with cognitive status, thus demonstrating the clinical validity of the features (Fig. 1).

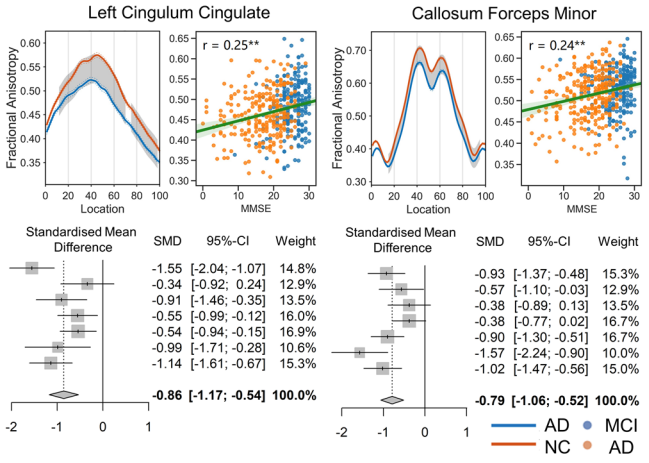

Figure 1. Significant variance in FA on left cingulum cingulate and callosum forceps minor between NC and AD. FA values along fibers and the correlation between the FA of the intervals with significant differences and MMSE were laid out. Forest plots showed the meta-analysis of the average FA values of significant locations along the fiber tract in AD and NC comparison. (**P<0.001)

From the meta-analysis, compared with NC, the AD group presented widely lower FA (10 tracts), fiber core linearity (12 tracts) and volume (8 tracts), and higher MD (18 tracts), RD (16 tracts) and AxD (14 tracts). The alteration patterns were similar but MD and RD were more sensitive to identify the alterations than other metrics. Bilateral cingulum, IFOF, uncinate fasciculus, and callosum forceps showed significant differences. The two fiber bundles that not only had a significant difference on all indicators but also had the largest effect size in FA were left cingulum cingulate (the largest Cohen's d in FA=-0.86, P=1.3e-07, shown as Fig 1 left) and callosum forceps minor (the second-largest Cohen's d in FA = -0.79, P=1.6e-08, shown as Fig. 1 right). However, in the left corticospinal, FA was significantly higher in AD group with a relatively lower effect size (Cohen's d=0.57, P=6.7e-06).

Correlation coefficients between the mean feature values of the locations with significant differences along significant tracts and MMSE showed commonly strong relevance (the absolute value of r ranging from 0.12 to 0.31). The strongest correlation averaged of all indicators was found on the callosum forceps minor (mean absolute r=0.25, r in FA=0.24, P<1e-04). The left cingulum cingulate also shown a strong correlation in FA (r=0.25, P<1e-04).

B. *Classification Results*

Based on the feature validity, the classification results demonstrated that the tract-based features had stable and generalizable classification performance.

When using original fiber features, the leave-one-site-out verification showed that the average ACC of 7 sites was 77.90%, 3 sites had ACC over 80%, the highest was 88.89% (S06), only one site had an ACC of under 70% (S02, 66.07%), and the average AUC was 0.86 (Table III), indicating validity and stability of tract-based characteristics in AD diagnosis. But the performances varied across sites, which may be caused by the different data distributions across sites (Fig. 2C).

TABLE III. Classification Performance

Index \ Site		S01	S02	S03	S04	S05	S06	S07	Average
Original data	*ACC*	82.76%	66.07%	75.00%	71.96%	84.13%	88.89%	76.47%	77.90%
	AUC	0.89	0.78	0.85	0.77	0.90	0.94	0.91	0.86
	SEN	84.09%	100.00%	62.50%	64.62%	76.60%	96.15%	92.86%	82.40%
	SPE	81.40%	26.92%	95.83%	83.33%	88.61%	78.95%	60.47%	73.64%
Harmonized data	*ACC*	81.61%	71.43%	79.69%	73.83%	87.30%	82.22%	82.35%	79.78%
	AUC	0.89	0.81	0.89	0.84	0.94	0.94	0.92	0.89
	SEN	81.82%	86.67%	75.00%	72.31%	85.11%	80.77%	90.48%	81.73%
	SPE	81.40%	53.85%	87.50%	76.19%	88.61%	84.21%	74.42%	78.02%

Correlation analysis showed that the decision function score in the AD group was negatively associated with MMSE (r=-0.32, P<1e-04, Fig. 2B). While as for the MCI group, the correlation coefficient was only -0.18 (P=0.005), manifesting that the classification output could represent the severity of cognitive impairment in the AD group but the predictive performance was limited in the MCI group.

Considering the site heterogeneity, we used the harmonized tract-based features processed via Combat to do the same experiment. Generally, data distribution differences between sites were reduced a lot (Fig. 2C). As Fig. 2D shown, the accuracies of sites with poor performance based on original data were greatly improved (S02, S03, S07) and the average performance was enhanced from ACC of 77.90% to 79.78% (AUC of 0.86 to 0.89). Besides, the performances of all sites were more balanced. The correlation between decision function and MMSE in AD group remained stable (r=-0.33).

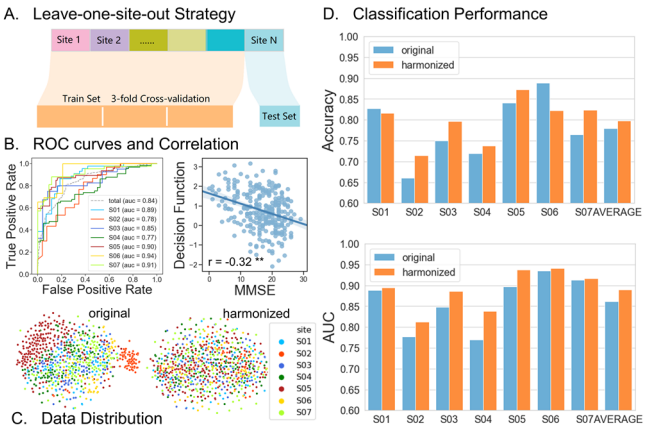

Figure 2. Classification performance. (A) leave-one-site-out verification strategy. (B-left) ROC curves of all sites in cross-validation with original features. (B-right) Correlations between classifier output (decision function with original features) and MMSE scores in the AD group. (C) The distribution of original data (left) and harmonized data (right) were shown. T-SNE (python, manifold-package) was used to visualize the latent space embedded by features. (D) Classification (SVM) performance comparison (ACC and AUC) between original feature and harmonized feature were shown.

Classification using XGBoost and linear kernel SVM showed similar performance (XGBoost: average AUC of original features = 0.85, average AUC of harmonized features = 0.87. linear SVM: average AUC of original features = 0.85, average AUC of harmonized features = 0.85). The tracts with higher average feature importance scores were left corticospinal, bilateral cingulum, callosum forceps, and uncinate fasciculus, enhancing the findings from the meta-analysis. We have also tried tri-classifications, but the results are limited with the best accuracy of only 55.20%. More detailed results can be found in [13].

IV. Discussion

Our study demonstrated the stability and generalizability of tract-based features extracted by AFQ in AD diagnosis, confirming the clinical utility of white matter parameters as biomarkers for AD. The advantage of our study is that, based on a large-scale multi-site dataset, the extensive lesions of WM fibers in AD were verified by meta-analysis, and the generalizability and stability of WM features in diagnosis were quantificationally measured by independent site cross-validation strategy, which made our results more credible and convincing.

From the perspective of the clinical significance of tract-based features, in line with previous studies [11, 12], WM alterations appeared widely in AD. And most of the changes were significantly correlated with the cognitive status in AD plus MCI groups, demonstrating a strong clinical basis of our findings. Among all the fiber tracts, the cingulum cingulate and callosum forceps have been confirmed to be most strongly associated with structural degeneration and impaired cognitive function in AD. Particularly, cingulum cingulate were found to be compromised in the early stage of AD [16]. The corpus callosum connects the two brain hemispheres and plays key roles in high-level associative connectivity [17, 18]. Hereby, meta-analysis enhanced the reliability of our results. Besides, the results of the meta-analysis are similar to the feature importance from classifiers, which can mutually verify.

Mining effective biomarkers play a vital role in the individualized treatment of AD. One of the great advantages of the present study is that the classifiers achieved high ACC with the leave-site-out independent cross-validation. This cross-validation strategy can not only reduce overfitting but also effectively quantify the generalizability of features and models across sites [14, 19]. According to the classification results, although the average performance was mediocre (ACC=77.90%), the accuracies of the top three were higher than 80%, illustrating that tract-based features were effective in AD diagnosis.

Amid the trend to use large-scale multisite datasets to boost statistical power, heterogeneity between sites is an inevitable problem for the reason of diverse acquisition conditions and processing pipelines (as shown in Fig. 2C). According to the classification results, the ACC of different sites was not balanced (the top was higher than 80% but the lowest was under 70%). The reasons for the site-dependent accuracies are complex. On the one hand, the accuracy of diagnosis (i.e., the real labels in classification) is one reason. Sites with accurate labels (e.g., S01, S06 collected by specialists), have higher classification accuracies. Improving the precision of labels is of great significance in building computer-aided diagnostic models. On the other hand, heterogeneity caused by the manufacturer may be one reason. The worst categorized site (S02) was the only one of the seven sites to use GE machines, and its data distribution deviated significantly from the other sites (Fig. 2C left), indicating that using data collected by different manufacturers requires more care. Among harmonization methods, ComBat has been confirmed to be an outstanding tool to reduce site heterogeneity meanwhile remain the biological information [15, 20]. In the classification experiment, the average ACC and AUC have been slightly improved and the accuracy of the site with poor performance has been enhanced after using ComBat, indicating the advantage of harmonization in classification. However, the disadvantage of ComBat is that it cannot deal with data from new sites which are very common in real applications [15]. In addition, the stability of performance improvement of the harmonization method among different algorithms remains to be further evaluated. Thus, how to combine the data from different sites and develop a computer-aided diagnostic model without site usage restriction is still a big challenge.

V. Conclusion

In conclusion, for the first time, we systematically assessed the validity and robustness of aberrant patterns of white matter dysconnectivity in AD using one of the largest DTI biobanks under a strict and unified analysis pipeline. The generalizability of tract-based features in the AD diagnosis was cross-validated with high prediction accuracies, proving the clinical practicality of white matter characteristics as biomarkers of AD.

Acknowledgment

Data were obtained from the Multi-Center Alzheimer's Disease Imaging Consortium. Data collection and sharing for this paper were funded by the National Natural Science Foundation of China (Nos. 61633018, 81571062, 81400890, 81471120, and 81701781).

References

[1] A. Burns and S. Iliffe, "Alzheimer's disease", *Bmj*, vol. 338, no. feb05 1, pp. b158-b158, 2009.

[2] J. Jia *et al.*, "The cost of Alzheimer's disease in China and re-estimation of costs worldwide", *Alzheimers Dement*, vol. 14, no. 4, pp. 483-491, 2018.

[3] B.P. Leifer, "Early diagnosis of Alzheimer's disease: clinical and economic benefits", *J Am Geriatr Soc*, vol. 51, no. 5 Suppl Dementia, pp. S281-288, 2003.

[4] J.M. Long and D.M. Holtzman, "Alzheimer Disease: An Update on Pathobiology and Treatment Strategies", *Cell*, vol. 179, no. 2, pp. 312-339, 2019.

[5] S. Rathore, M. Habes, M.A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages", *Neuroimage*, vol. 155, pp. 530-548, 2017.

[6] Y. Jin *et al.*, "3D tract-specific local and global analysis of white matter integrity in Alzheimer's disease", *Human Brain Mapping*, vol. 38, no. 3, pp. 1191-1207, 2017.

[7] J.D. Yeatman, R.F. Dougherty, N.J. Myall, B.A. Wandell, and H.M. Feldman, "Tract profiles of white matter properties: automating fiber-tract quantification", *PLoS One*, vol. 7, no. 11, pp. e49790, 2012.

[8] J.D. Yeatman, A. Richie-Halford, J.K. Smith, A. Keshavan, and A. Rokem, "A browser-based tool for visualization and analysis of diffusion MRI data", *Nat Commun*, vol. 9, no. 1, pp. 940, 2018.

[9] M.D. Sacchet *et al.*, "Structural abnormality of the corticospinal tract in major depressive disorder", *Biology of Mood & Anxiety Disorders*, vol. 4, no. 1, pp. 8, 2014.

[10] H. Sun *et al.*, "Two Patterns of White Matter Abnormalities in Medication-Naive Patients With First-Episode Schizophrenia Revealed by Diffusion Tensor Imaging and Cluster Analysis", *JAMA Psychiatry*, vol. 72, no. 7, pp. 678-686, 2015.

[11] X. Zhang *et al.*, "Characterization of white matter changes along fibers by automated fiber quantification in the early stages of Alzheimer's disease", *NeuroImage: Clinical*, vol. 22, pp. 101723, 2019.

[12] X. Dou *et al.*, "Characterizing white matter connectivity in Alzheimer's disease and mild cognitive impairment: An automated fiber quantification analysis with two independent datasets", *Cortex*, vol. 129, pp. 390-405, 2020.

[13] Y. Qu *et al.*, "AI4AD: Artificial intelligence analysis for Alzheimer's disease classification based on a multisite DTI database", *Brain Disorders*, vol. 1, pp. 100005, 2021.

[14] D. Jin *et al.*, "Generalizable, Reproducible, and Neuroscientifically Interpretable Imaging Biomarkers for Alzheimer's Disease", *Adv Sci*, vol. 7, no. 14, pp. 2000675, 2020.

[15] J.P. Fortin *et al.*, "Harmonization of multi-site diffusion tensor imaging data", *Neuroimage*, vol. 161, pp. 149-170, 2017.

[16] M. Catani, R.J. Howard, S. Pajevic, and D.K. Jones, "Virtual in vivo interactive dissection of white matter fasciculi in the human brain", *Neuroimage*, vol. 17, no. 1, pp. 77-94, 2002.

[17] M.G. Preti *et al.*, "Assessing Corpus Callosum Changes in Alzheimer's Disease: Comparison between Tract-Based Spatial Statistics and Atlas-Based Tractography", *PLOS ONE*, vol. 7, no. 4, pp. e35856-e35856, 2012.

[18] H. Hanyu, T. Asano, H. Sakurai, Y. Imon, and K. Abe, "Diffusion-weighted and magnetization transfer imaging of the corpus callosum in Alzheimer's disease", *Journal of the Neurological Sciences*, vol. 167, no. 1, pp. 37-44, 1999.

[19] C. Davatzikos, "Machine learning in neuroimaging: Progress and challenges", *Neuroimage*, vol. 197, pp. 652-656, 2019.

[20] M. Yu *et al.*, "Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data", *Human Brain Mapping*, vol. 39, pp. 4213-4227, 2018.