Early Diagnosis of Alzheimer through Swin-Transformer-Based Deep Learning Framework using Sparse Diffusion Measures

Abhishek Tiwari* AT326@SNU.EDU.IN
Ananya Singhal* AS146@SNU.EDU.IN
Saurabh J. Shigwan* SAURABH.SHIGWAN@SNU.EDU.IN
Rajeev Kumar Singh* RAJEEV.KUMAR@SNU.EDU.IN
*for the Alzheimer's Disease Neuroimaging Initiative***
**Shiv Nadar Institution of Eminence, Delhi NCR, India*

Editors: Berrin Yanıkoğlu and Wray Buntine

Abstract

Alzheimer disease is one of the most common neuro-degenerative diseases, with an estimated 6.2 million cases in the United States. This research article investigates the potential of Transformer-based deep learning techniques to accelerate the processing of diffusion tensor imaging (DTI) measures and improve the early diagnosis of Alzheimer disease (AD) using sparse data. Diffusion Weighted Imaging (DWI) is a time-consuming process, with each diffusion direction taking between 2-5 minutes, and at least 40 diffusion directions are needed for routine clinical diagnosis, which needs scanning duration exceeding 3 hours for each patient. By leveraging the attention mechanism, our proposed model generates quantitative measures of fractional anisotropy (FA), axial diffusivity (AxD), and mean diffusivity (MD) using 5 and 21 diffusion directions, making it useful for clinical diagnosis through reduced scanning time of more than half. Our experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate that our proposed model outperforms the traditional linear least square method, achieving accurate quantitative measurement of FA, AxD, and MD scores for early diagnosis of AD patients from healthy controls using sparse diffusion directions. Our analysis highlights the potential of Swin-Transformer attention-based deep learning framework to improve the early diagnosis and treatment of Alzheimer's disease. A repositories for our research work at https://github.com/AbhishekTiwari101/ACML2023-Early-Diagnosis-of-Alzheimer-via-Deep-Learning https://github.com/reachananya/Early-diagnosis-of-Alzheimer-via-DL

**Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:ADNI Acknowledgement List

Keywords: Alzheimer diseases, Diagnosis, Deep learning, Sparse diffusion measures, DTI.

1. Introduction

Millions of people worldwide are affected by Alzheimer disease (AD), which is a type of neurodegenerative disorder Kolahkaj et al. (2023); Qu et al. (2021). Routine clinical diagnosis of AD involves DWI, which is a time-consuming process, with each diffusion direction taking between 2-5 minutes, and at least 40 diffusion directions are needed. Consequently, each patient's scanning duration exceeds three hours Scherrer and Warfield (2012). The effective treatment and management of AD rely heavily on early and accurate diagnosis Marzban et al. (2020); Gao and Lima (2022). A powerful neuroimaging technique called diffusion tensor imaging (DTI) can provide valuable information on the microstructural integrity of the brain's white matter, which is often altered in Mild Cognitive Impairment (MCI), an early stage of AD Nir et al. (2013); Zhan et al. (2014); Demirhan et al. (2015). However, the traditional linear least square method used for DTI processing requires a large number of diffusion directions, making it challenging for both the patient and the diagnostic lab.

Recent advances in deep learning Aja-Fernández et al. (2023); Liu et al. (2022); Tiwari and Singh (2022); Aderghal et al. (2020) have shown tremendous potential in accelerating the processing of DTI measures and improving the quantitative measure of MCI and AD diagnosis. Attention-based Karimi and Gholipour (2022) deep learning is a powerful technique that can focuses on relevant regions of interest in the brain and extracts diffusion tensor imaging quantitative measures viz fractional anisotropy (FA), axial diffusivity (AxD), and mean diffusivity (MD). By leveraging the attention mechanism Karimi and Gholipour (2022), it is possible to process DTI measures more efficiently and accurately, even with six diffusion directions. The utilization of the multi-head self-attention mechanism, as introduced in Vaswani et al. work on attention Vaswani et al. (2017), is employed by Transformer-DTI Karimi and Gholipour (2022) to leverage the spatial correlation present in diffusion tensor parameters and diffusion signals across adjacent slices and voxels. Despite the immense promise of transformer models, their scalability becomes a concern when dealing with a larger number of diffusion directions. This is primarily due to the increase in trainable parameters, resulting in longer training times and higher memory requirements. Consequently, training a single model capable of handling multiple diffusion directional signals becomes a challenging task. However, the Swin Transformer model has garnered significant attention for its exceptional ability to handle larger input sizes, setting it apart from traditional transformer models Liu et al. (2021). This deep neural network architecture has recently emerged and demonstrated impressive performance across various computer vision tasks, such as semantic segmentation, object detection, and image classification. Unlike conventional transformer models that process input data sequentially Vaswani et al. (2017), the Swin Transformer adopts a hierarchical structure.

In the Swin Transformer, input feature maps are divided into non-overlapping patches, and these patches are simultaneously processed using multiple transformer layers. This approach enables efficient parallelization and reduces the computational cost of the self-attention mechanism. Consequently, the number of parameters is reduced, which is crucial for developing a more generalized model capable of handling multiple diffusion directional signals. In the context, we propose a novel Swin-Transformer attention-based deep learning model that can extract quantitative measures such as FA, AxD, and MD, to accurately diagnose early incidence of AD. To evaluate the effectiveness of our proposed model, we

employ the openly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. We compare our model performance with a traditional linear least square method Koay et al. (2006) for 41 diffusion directions. Furthermore, for 21 and 5 diffusion directions, we compare our model with both the traditional linear least square method Koay et al. (2006) and Transformer-DTI Karimi and Gholipour (2022). Through these comparisons, we aim to demonstrate the superiority of our proposed model in accurately diagnosing AD at its early stage.

The main contributions of this article are:

- Efficient learning of spatial correlation in neighboring voxels to enhance robustness of diffusion tensor imaging parameters through a unique training strategy involving shifting window with overlapping strides.
- Accelerated estimation of DTI parameters using sparse measurements facilitated by the proposed method.
- Demonstrated capability to quantify diffusion parameters and detect early-stage Alzheimer's disease using the proposed model.

The rest of the paper is organized as follows. In Section 2, we describe our proposed method and its architecture in detail. In Section 3, we present the experimental results and analysis, and in Section 4, we conclude the paper with a summary of our findings and future research directions.

2. METHOD

A DWI image is characterized by four dimensions, where the first three dimensions correspond to spatial measurements, and the fourth dimension represents the measurement of water diffusivity across multiple diffusion directions. DTI being specific type of DWI, can capture diffusion anisotropy, by considering diffusivity as a function of orientation through a Gaussian model Le Bihan et al. (2001); Basser et al. (1994). This orientation function is characterized with 3×3 symmetric matrix D as represented in equation 1.

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \tag{1}$$

The tensor D represents the magnitude of diffusivity on an ellipsoid's surface. Its eigenvectors and eigenvalues correspond to the directions and lengths of the ellipsoid's axes, respectively. Diffusion signal inside each 3D voxel is a N dimensional vector represented as $\mathbf{s} = [s_1, s_2, ..., s_N]$. According to DTI formulation Le Bihan et al. (2001), as represented in equation 2, we want to extract diffusion tensor D from signal $\mathbf{s}$.

$$s_i = s_0 e^{(-b_{xx}D_{xx} - b_{yy}D_{yy} - b_{zz}D_{zz} - 2b_{xy}D_{xy} - 2b_{xz}D_{xz} - 2b_{yz}D_{yz})} \tag{2}$$

Here, b_{jk}, for $j, k = x, y, z$, represents function of diffusion directions $\mathbf{g} = \{g_i\}_{i=1}^{N}$, where each $g_i = [g_{ix}, g_{iy}, g_{iz}]$ is a unit direction vector.Along with that s_0 represents zero diffusion signal. Using simplified assumption of the b-matrix as given in Le Bihan et al.

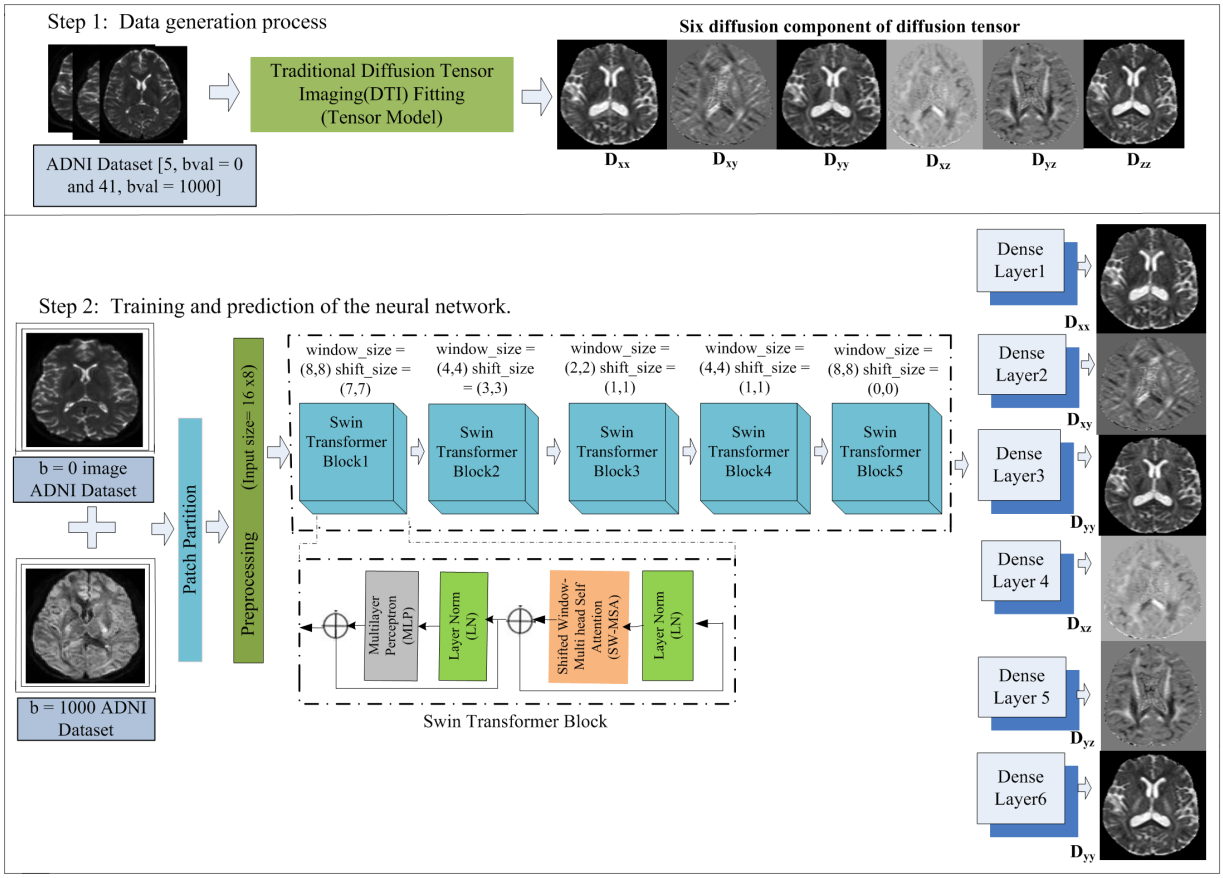

Figure 1: The proposed model architecture for training and prediction of the neural network.

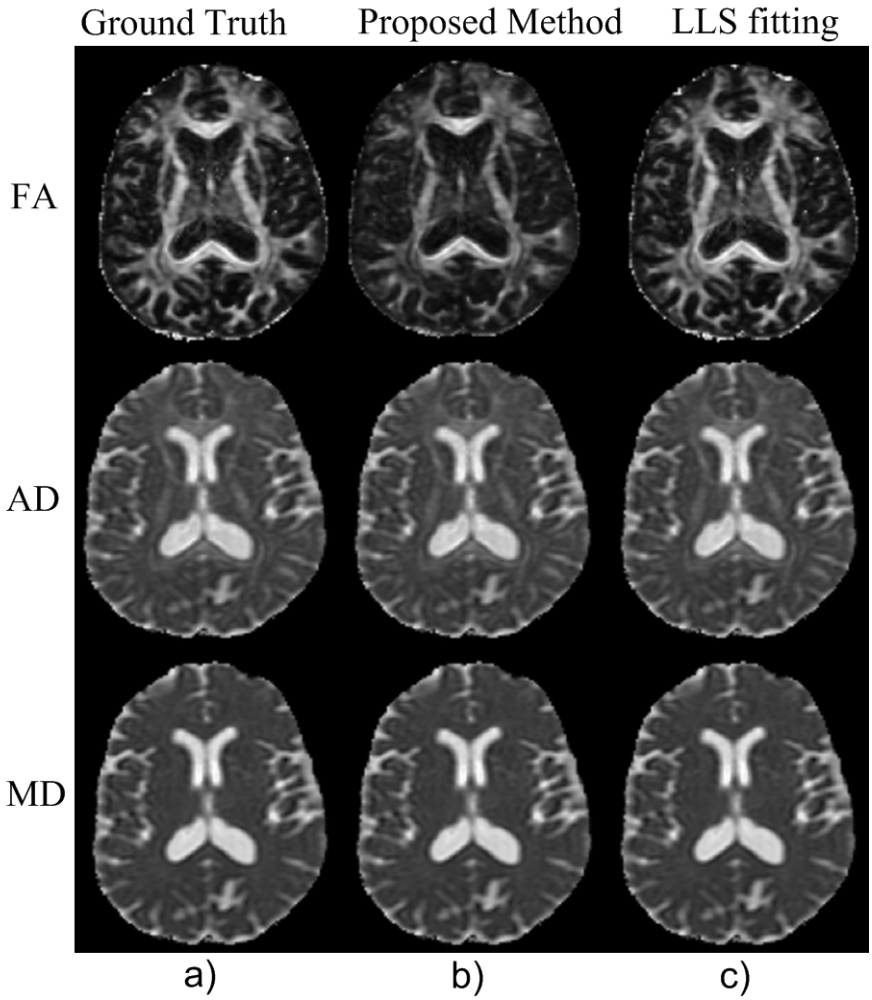

Figure 2: Comparing outcomes from ground truth, the proposed method, and LLS fitting Koay et al. (2006), we note that the Proposed method demonstrates comparability with both the ground truth and LLS fitting Koay et al. (2006) for 41 diffusion directions.

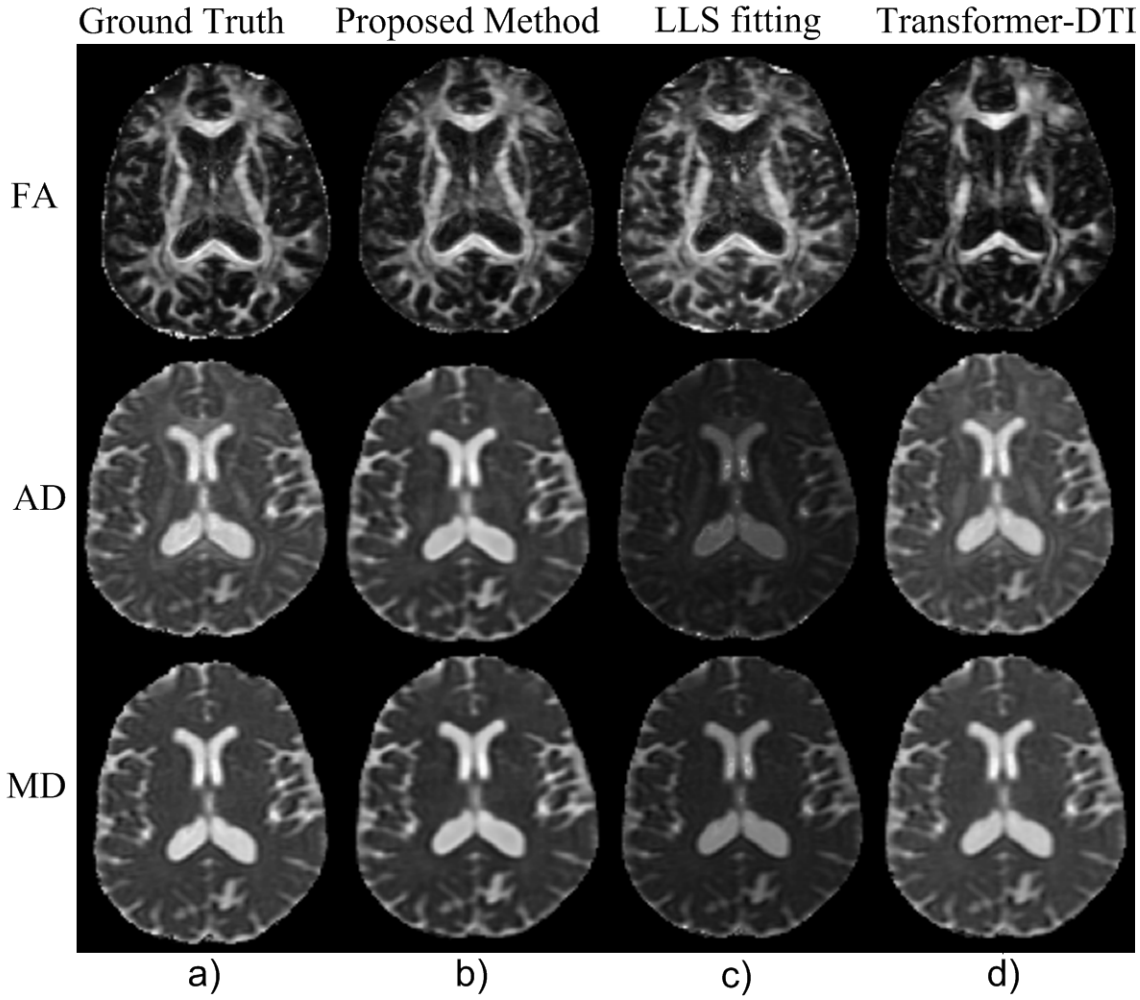

Figure 3: Comparing outcomes from ground truth, the proposed method, LLS fitting Koay et al. (2006) and Transformer-DTI Karimi and Gholipour (2022), we note that the Proposed method demonstrates comparability with the ground truth and surpasses both LLS fitting Koay et al. (2006) and Transformer-DTI Karimi and Gholipour (2022) for 21 diffusion directions.

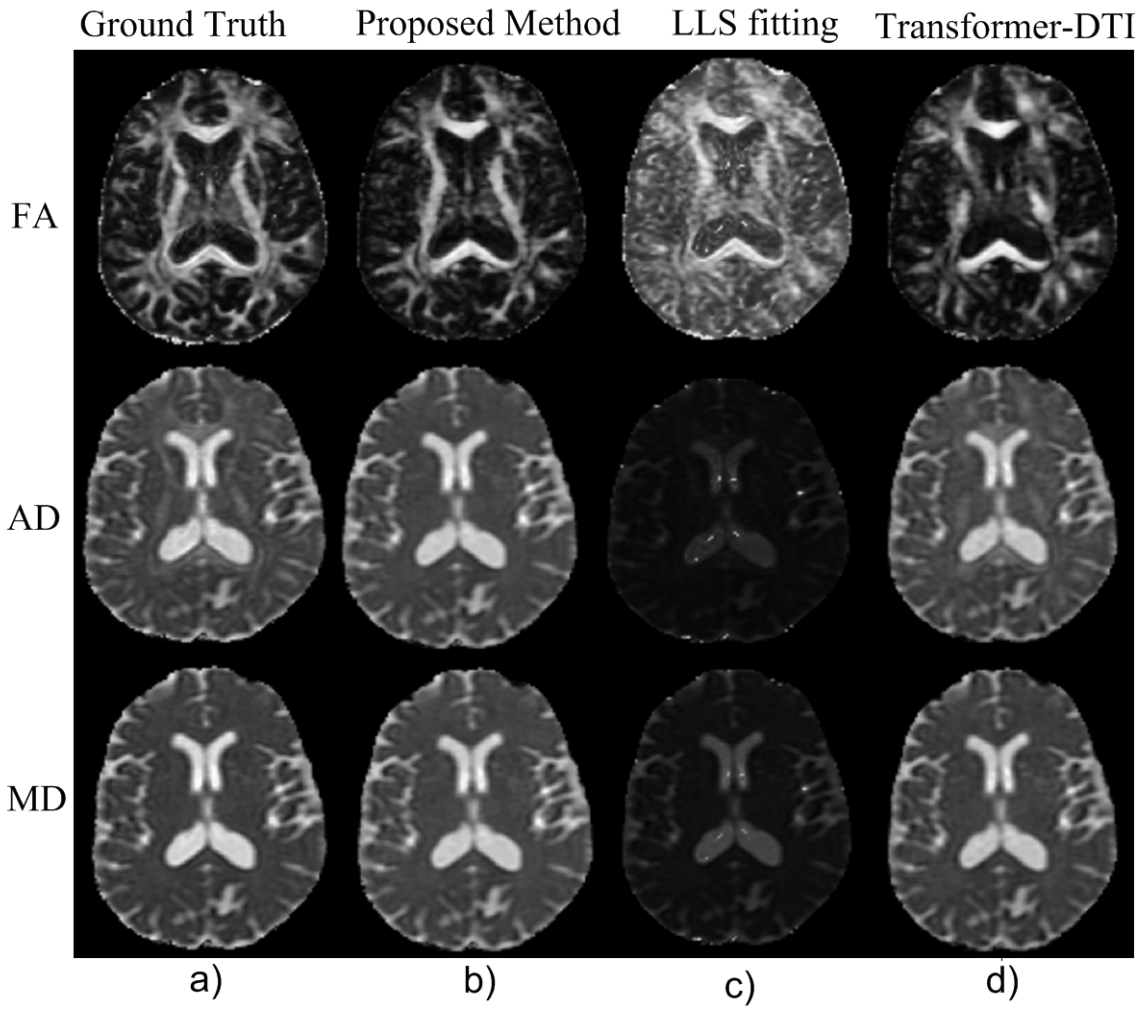

Figure 4: Comparing outcomes from ground truth, the proposed method, LLS fitting Koay et al. (2006), and Transformer-DTI Karimi and Gholipour (2022), we observe that the Proposed method achieves comparability with the ground truth and surpasses LLS fitting Koay et al. (2006) and Transformer-DTI Karimi and Gholipour (2022) for 5 diffusion directions.

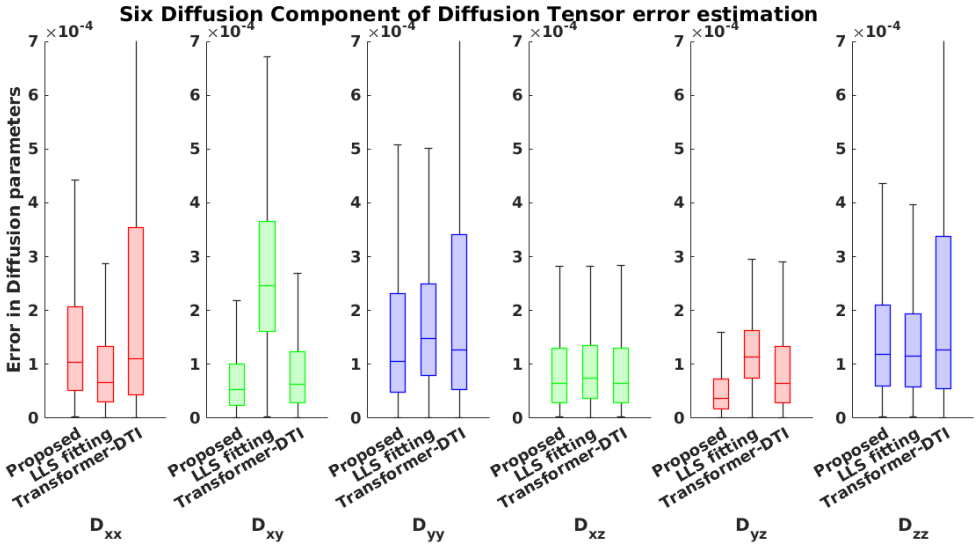

Figure 5: Six diffusion component of diffusion tensor for 5 diffusion directions

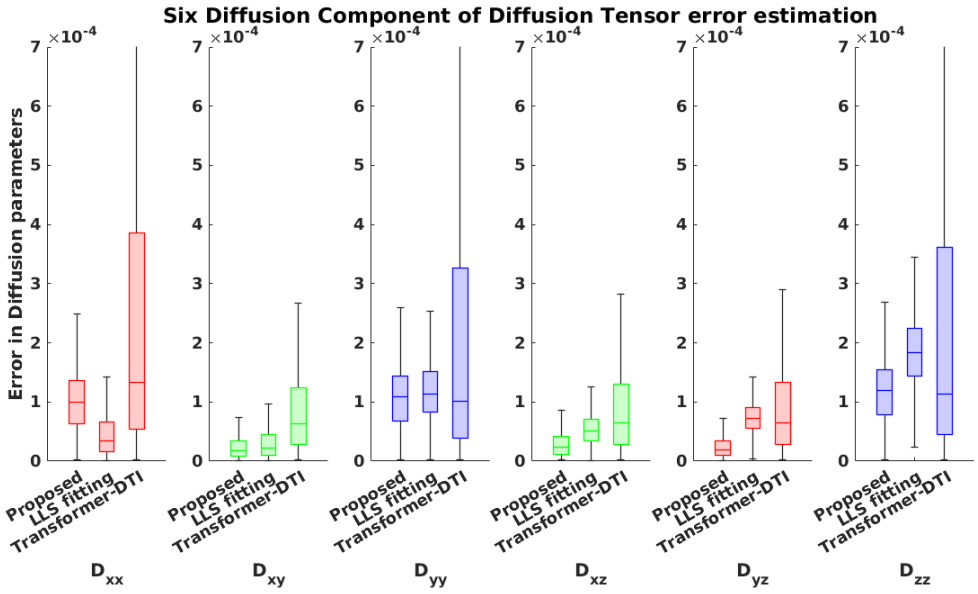

Figure 6: Six diffusion component of diffusion tensor for 21 diffusion directions

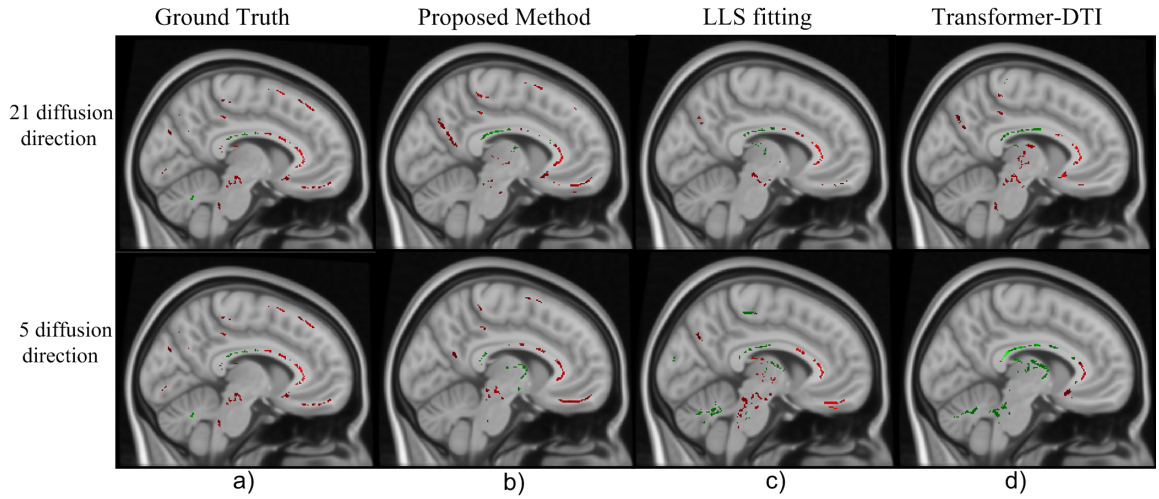

Figure 7: Axial brain slice representing the Cingulum region, highlighted with p-value of two sample t-test(df=22). Green color: hypothesis testing tstat1 - Healthy CN > MCI, Red color: hypothesis testing tstat2 - Healthy CN < MCI.

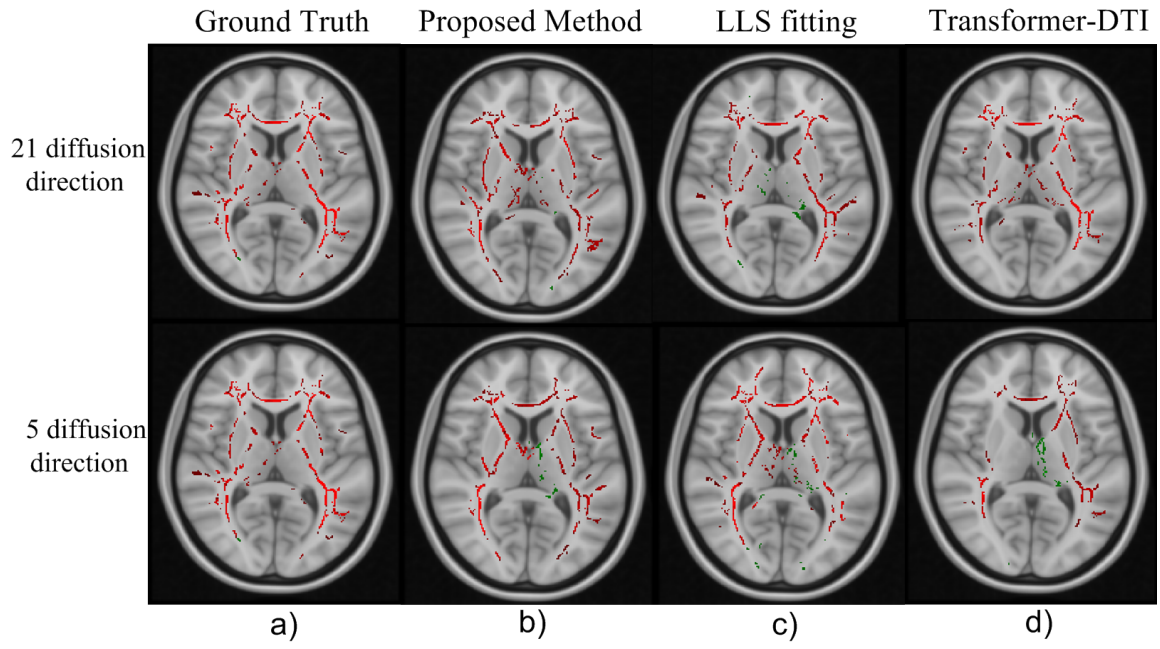

Figure 8: Coronal brain slice representing the Uncinate fasciculus region, highlighted with p-value of two sample t-test(df=22). Green color: hypothesis testing tstat1 - Healthy CN > MCI, Red color: hypothesis testing tstat2 - Healthy CN < MCI.

(2001), we obtained equation $s_i/s_0 = e^{-bg_i^T Dg_i}$ which provides best accuracy of diffusion coefficient D. The apparent diffusion coefficient K_i with respect to each g_i can be calculated as $K_i = g_i^T Dg_i = (-1/b)\ln(s_i/s_0)$. The design matrix $\alpha = [\alpha_1, ..., \alpha_N]^T$ represents linear mapping between diffusion tensor D and apparent diffusion coefficient estimates $K = [K_1, ..., K_N]^T$ using $K = \alpha\bar{D}$, where $\bar{D} = [D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, D_{yz}]^T$ and $\alpha_i = [g_{ix}^2, g_{iy}^2, g_{iz}^2, 2g_{ix}g_{iy}, 2g_{ix}g_{iz}, 2g_{iy}g_{iz}]^T$. Traditional linear least fitting (LLS) Koay et al. (2006) method uses $\bar{D} = (\alpha^T\alpha)^{-1}\alpha^T K$, which is susceptible to noisy and sparse measurements. For sparse measurements, instead of solving using linear model, we proposed to solve for D directly using equation 2 as inverse map $\bar{D} = F(\mathbf{X}, \mathbf{g})$, where input $\mathbf{X} = [s_1/s_0, ..., s_N/s_0]$ per voxel. We have formulated our F using Swin-transformer based neural network Liu et al. (2021).Our modeling framework is more generic framework, where single model is sufficient to learn from diffusion signal with different number of diffusion directions. We have experimented with $\{41, 21, 5\}$ number of diffusion directions. In practical applications, multiple attention heads are commonly used to learn diverse representations of the input. To capture pixel relationships in the vicinity, a softmax function is applied to the logits computed in that particular neighborhood. The Transformer-DTI Karimi and Gholipour (2022) model employed multi-head self-attention to estimate diffusion tensor imaging parameters using only six diffusion-weighted images.

In our proposed model, we introduce a modified version of window self-attention to enhance non-linearity and decrease the number of parameters. The window self-attention mechanism in our model offers several advantages over traditional self-attention mechanisms, making it a crucial component. Instead of integrating the diffusion direction using learned weights, we concatenate the direction vector as part of the input signal. This enhancement improves the model's non-linearity and enables it to capture more intricate relationships within the input data. Traditionally, self-attention applies distinct triplets of (query, key, value) to each position in the input. However, in our proposed model, we employ window self-attention, where the same triplets are applied to all windows within the same patch. This patch-based method, divides the input image into patches and independently applies spatial relative attention to each patch. This strategy reduces the computational burden of spatial relative attention and allows for the modeling of long-range spatial relationships in large images.

2.1. Data Preprocessing and Model Training

We developed an algorithm to preprocess the DWI images from the ADNI dataset, creating three preprocessed training datasets: $ADNI-41$, $ADNI-21$, and $ADNI-5$. The algorithm is outlined in Algorithm 1.

Our neural network was trained on a dataset consisting of 1 million tuples per image obtained from three different datasets: $ADNI-41$, $ADNI-21$, and $ADNI-5$. These tuples were derived from a subset of 10 DWI images, which were pre-processed using *Algorithm 1*. The architecture of our neural network is illustrated in Figure 1, where we employ a Swin-transformer block known for its ability to capture correlations within input signals. In our case, the input signals consist of diffusion signals along with diffusion directions, and our model considers the diffusion signals within a neighboring $5 \times 5 \times 5$ patch along with the current voxel diffusion signal.

Algorithm 1 Data preprocessing of proposed model

Output: Algorithm for preprocessing DWI images of ADNI dataset, creating preprocessed training datasets $ADNI-41$, $ADNI-21$, and $ADNI-5$.

Step 1: Apply the DTI model fit from the DIPY Python package Garyfallidis et al. (2014) to each DWI image individually, resulting in six diffusion components: D_{xx}, D_{xy}, D_{yy}, D_{xz}, D_{yz}, and D_{zz} of the diffusion tensor.
Step 2: Select $100,000$ voxels from each DWI image based on their fractional anisotropy (FA) score, excluding zero, uniformly distributed within the range $(0, 1)$.
Step 3: Obtain $100,000$ tuples of (input, ground-truth), where each tuple consists of an input $5 \times 5 \times 5$ voxel patch with 41 diffusion directions per voxel, and the ground-truth is a $5 \times 5 \times 5$ voxel patch with a 6×1 vector per voxel, representing the corresponding six diffusion components of the diffusion tensor.
Step 4: Concatenate the diffusion signals of the neighboring $5 \times 5 \times 5$ patch to form a 125×41 matrix per input voxel.
Step 5: Concatenate the diffusion directions (3×41) with the diffusion signal to obtain a 128×41 matrix per input voxel.
Step 6: Zero-pad the input of each tuple to create a 128×100 matrix, while the ground truth is represented as a 125×6 matrix. This resulting training dataset is denoted as $ADNI-41$.
Step 7: Perform Qball-based interpolation Tuch (2004); Garyfallidis et al. (2014) on the 41-directional input of the $ADNI-41$ dataset to obtain 41-directional, 21-directional, and 5-directional diffusion signals, using the DIPY package.
Step 8: Concatenate the diffusion directions with the diffusion signals of each tuple to create inputs of size 125×41, 125×21, and 125×5 vectors, and a ground truth of size 125×6 matrix.
Step 9: Zero-pad the input of each tuple to create a 128×100 matrix.
Step 10: The resulting training datasets are referred to as $ADNI-41$, $ADNI-21$, and $ADNI-5$, corresponding to the respective diffusion directions $= 41, 21, 5$.

3. Results and Discussions

For our experiments, we utilized 40 DWI images of individuals in the ADNI dataset. Among these images, 20 are from the Cognitively Normal (CN) group and the remaining 20 are from the Mild Cognitive Impairment (MCI) group. To train our model, we selected 5 CN and 5 MCI images, while 3 CN and 3 MCI images were used for validation. Finally, we reserved 12 CN and 12 MCI images for testing. For our experiments, we have considered the output of LLS fitting over $ADNI - 41$ as ground truth. To assess the performance of our proposed method, we compared the results with LLS fitting and Transformer-DTI. Figure 2 illustrates comparisons of ground truth, proposed method, and LLS fitting for 41 diffusion directions. The results demonstrate that the proposed method is equally comparable LLS fitting. These findings indicate that the proposed method provides an accurate estimation of diffusion tensor. The results presented in figure 3 compare the diffusion measures of 21 diffusion directional signal using proposed method, LLS fitting and Transformer-DTI. The results presented in figure 4 of compare the diffusion measure of 5 diffusion directional signal using proposed method, LLS fitting, and Transformer-DTI. Notable, our proposed method demonstrates comparable performance to LLS fitting Koay et al. (2006) and outperforms Transformer-DTI Karimi and Gholipour (2022) for diffusion directions of 21 as mentioned in figures 3 and 6. For 5 diffusion directional, our proposed model outperforms both LLS fitting Koay et al. (2006) and Transformer-DTI Karimi and Gholipour (2022), as per mentioned figures 4 and 5. These results highlight the effectiveness of our method in accurately estimating diffusion measures, particularly for a lower number of diffusion directions. The results presented in figures 5 and 6 demonstrates the error plots of six diffusion components $\bar{D}$ for 5 and 21 diffusion directional signal, respectively.

To further assess the statistical significance of our results, we conducted a t-test analysis using tract-based spatial statistics (TBSS) pipeline Smith et al. (2006) to perform a statistical analysis comparing two groups (Healthy CN and MCI) based on fractional anisotropy (FA) metrics. FA images were non-linearly registered to the FMRIB-58 template in the Montreal Neurological Institute (MNI) space, which includes averaged FA maps, using the FNIRT tool from FSL Rueckert et al. (1999). The white matter skeleton was defined by thinning a mean FA image generated with an FA threshold of 0.2 to differentiate white matter from gray matter. This rigorous approach allowed for reliable statistical interpretation of the data and facilitated a comprehensive comparison of the two groups. In figure 7, an axial brain slice representing the Cingulum region is highlighted with a p-value obtained from a two sample t-test with 22 degrees of freedom. Similarly, in Figure 8, a coronal brain slice representing the Uncinate fasciculus region is highlighted with a p-value obtained from a two sample t-test with 22 degrees of freedom. These figures provide valuable information about the regions of the brain that may be affected by the experimental conditions and can be further discussed in the context of the research question or hypothesis of the study.

The results of the studies conducted by Fani et al. (2014); Duncan (2010) demonstrate a significant association between the development of Alzheimer's disease and the Cingulum and Uncinate fasciculus. Figure 7 and figure 8 demonstrate similar relationship using our proposed framework with sparse data, thereby reducing scanning time significantly. The Cingulum and Uncinate fasciculus are essential white matter tracts in the brain that play a critical role in diverse cognitive processes, including memory, emotional control, and social

Table 1: Number of pixels with p-Value outside Confidence Intervals (95% and 99%) for t-statistics (tstat1 and tstat2) in Proposed Model, LLS fitting, and Transformer-DTI. Hypothesis tstat1: Healthy CN > MCI, hypothesis tstat2: Healthy CN < MCI.

P-Value	**Ground Truth**	**Proposed Model**		**LLS fitting**		**Transformer-DTI**	
		5 Diff.	**21 Diff.**	**5 Diff.**	**21 Diff.**	**5 Diff.**	**21 Diff.**
tstat1, 95 C.I.	3775	3246	3146	4024	5568	5311	4646
tstat2, 95 C.I.	7586	8154	7188	7836	5683	5259	7228
tstat1, 99 C.I.	618	505	485	741	981	960	860
tstat2, 99 C.I.	1352	1693	1330	1578	917	1012	1184

behavior. Recent research suggests that modifications in the microstructural integrity of these tracts may be linked to the onset of Alzheimer's disease. Table 1 provides number of pixels with p-values outside confidence intervals. It exhibits the outcomes for different diffusion directions, namely 5 Diff. and 21 Diff. It provides confidence intervals at both 95% (95 C.I.) and 99% (99 C.I.) confidence levels. The results in Table 1 demonstrate that the proposed method preserves an equal number of pixels with significant differences in two groups compared to the ground truth. In contrast, the LLS fitting and Transformer-DTI methods yield a number of pixels with significant differences in two groups, deviating further from the ground truth.

4. CONCLUSION

In conclusion, the proposed Swin-Transformer-based deep learning framework that incorporates sparse diffusion measures is a promising approach for early diagnosis of Alzheimer disease. The use of sparse diffusion measures in the proposed framework provided an effective way to capture the underlying structural connectivity of the brain, which is a key feature in Alzheimer disease. In our proposed model, built upon the foundation of Swin Transformer architecture, we introduced a strategic manipulation of neighboring windows through a designated stride. This deliberate adjustment resulted in regions of overlap between adjacent windows. This intentional overlapping had the specific purpose of enhancing inter-token attention. This enhancement allowed tokens to establish meaningful connections even when they exist within separate windows. This dynamic interplay between tokens introduced a broader range of possibilities for mutual influence, thus fostering the creation of more intricate and complex representations.

The Swin-Transformer-based deep learning framework was able to effectively learn the neighboring patterns which inherently represent complex relationships between brain connectivity and disease status, thereby enabling accurate diagnosis of Alzheimer disease at an early stage. Further research can be done to explore the potential of this framework in large-scale clinical trials and to investigate its effectiveness in detecting Alzheimer disease in different populations.

Acknowledgments

We would like to thank Shiv Nadar University, Delhi NCR, India, for the support throughout the research process. We are also grateful to the reviewers and editor, whose valuable feedback and suggestions helped in improving the quality of the manuscript.

References

Karim Aderghal, Karim Afdel, Jenny Benois-Pineau, and Gwénaëlle Catheline. Improving alzheimer's stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities. *Heliyon*, 6(12), 2020.

Santiago Aja-Fernández, Carmen Martín-Martín, Álvaro Planchuelo-Gómez, Abrar Faiyaz, Md Nasir Uddin, Giovanni Schifitto, Abhishek Tiwari, Saurabh J Shigwan, Rajeev Kumar Singh, Tianshu Zheng, et al. Validation of deep learning techniques for quality augmentation in diffusion mri for clinical studies. *NeuroImage: Clinical*, page 103483, 2023.

Peter J Basser, James Mattiello, and Denis LeBihan. Mr diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.

Ayşe Demirhan, Talia M Nir, Artemis Zavaliangos-Petropulu, Clifford R Jack, Michael W Weiner, Matt A Bernstein, Paul M Thompson, and Neda Jahanshad. Feature selection improves the accuracy of classifying alzheimer disease using diffusion tensor images. pages 126–130, 2015.

John S Duncan. Imaging in the surgical treatment of epilepsy. *Nature Reviews Neurology*, 6(10):537–550, 2010.

Negar Fani, Tricia Z King, Emily Reiser, Elisabeth B Binder, Tanja Jovanovic, Bekh Bradley, and Kerry J Ressler. Fkbp5 genotype and structural integrity of the posterior cingulum. *Neuropsychopharmacology*, 39(5):1206–1213, 2014.

Shuangshuang Gao and Dimas Lima. A review of the application of deep learning in the detection of alzheimer's disease. *International Journal of Cognitive Computing in Engineering*, 3:1–8, 2022.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, Ian Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

Davood Karimi and Ali Gholipour. Diffusion tensor estimation with transformer neural networks. *Artificial Intelligence in Medicine*, page 102330, 2022.

Cheng Guan Koay, Lin-Ching Chang, John D Carew, Carlo Pierpaoli, and Peter J Basser. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of magnetic resonance*, 182(1):115–125, 2006.

Shayan Kolahkaj et al. A connectome-based deep learning approach for early mci and mci detection using structural brain networks. *Neuroscience Informatics*, page 100118, 2023.

Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 13(4):534–546, 2001.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. pages 10012–10022, 2021.

Zhenbing Liu, Haoxiang Lu, Xipeng Pan, Mingchang Xu, Rushi Lan, and Xiaonan Luo. Diagnosis of alzheimer's disease via an attention-based multi-scale convolutional neural network. *Knowledge-Based Systems*, 238:107942, 2022.

Eman N Marzban, Ayman M Eldeib, Inas A Yassine, Yasser M Kadah, and Alzheimer's Disease Neurodegenerative Initiative. Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks. *PloS one*, 15(3):e0230409, 2020.

Talia M Nir, Neda Jahanshad, Julio E Villalon-Reina, Arthur W Toga, Clifford R Jack, Michael W Weiner, Paul M Thompson, Alzheimer's Disease Neuroimaging Initiative (ADNI, et al. Effectiveness of regional dti measures in distinguishing alzheimer's disease, mci, and normal aging. *NeuroImage: clinical*, 3:180–195, 2013.

Yida Qu, Pan Wang, Bing Liu, Chengyuan Song, Dawei Wang, Hongwei Yang, Zengqiang Zhang, Pindong Chen, Xiaopeng Kang, Kai Du, et al. Ai4ad: artificial intelligence analysis for alzheimer's disease classification based on a multisite dti database. *Brain Disorders*, 1:100005, 2021.

Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

Benoit Scherrer and Simon K Warfield. Parametric representation of multiple white matter fascicles from cube and sphere diffusion mri. *PLoS one*, 7(11):e48232, 2012.

Stephen M Smith, Mark Jenkinson, Heidi Johansen-Berg, Daniel Rueckert, Thomas E Nichols, Clare E Mackay, Kate E Watkins, Olga Ciccarelli, M Zaheer Cader, Paul M Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

Abhishek Tiwari and Rajeev Kumar Singh. Performance, trust, or both? covid-19 diagnosis and prognosis using deep ensemble transfer learning on x-ray images. pages 1–9, 2022.

David S Tuch. Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(6):1358–1372, 2004.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Liang Zhan, Matt A Bernstein, B Borowski, Clifford R Jack, and Paul M Thompson. Evaluation of diffusion imaging protocols for the alzheimer's disease neuroimaging initiative. pages 710–713, 2014.