Evaluating the Stability of Deep Neural Networks for Brain Microstructural Parameter Estimation under Varying Diffusion Gradient Schemes

Dominika Ciupek[1], Maciej Malawski[1,2], Tomasz Pięciak[3,1]

[1]Sano Centre for Computational Medicine, Krakow, Poland, [2]AGH University of Krakow, Krakow, Poland, [3]Universidad de Valladolid, Valladolid, Spain

INTRODUCTION

- Bipolar disorder (BD) is a chronic psychiatric condition characterized by significant fluctuations in mood and energy, with diagnosis often delayed by 5–10 years [1].
- Diffusion-weighted magnetic resonance imaging (DW MRI) offers promise for identifying objective biomarkers in psychiatric assessment. In particular, diffusion tensor imaging (DTI) [2] enables the analysis of brain microstructure, which may help distinguish BD patients from healthy controls (HC).
- However, standard acquisition protocols and evaluation methods remain insufficient for ensuring clinically meaningful accuracy.

OBJECTIVE OF THE STUDY

This study aims to evaluate the potential of deep learning (DL) models to accurately estimate DTI microstructural parameters, while preserving clinically relevant differences between BD patients and HC.

MATERIALS AND METHODS

DTI: Two microstructural parameters were estimated:

1) radial diffusivity (RD): 2) fractional anisotropy (FA):

$$\mathrm{RD} = \frac{\lambda_2 + \lambda_3}{3};\qquad \mathrm{FA} = \sqrt{\frac{1}{2}}\frac{\sqrt{(\lambda_1-\lambda_2)^2 + (\lambda_1-\lambda_3)^2 + (\lambda_2-\lambda_3)^2}}{\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2}}.$$

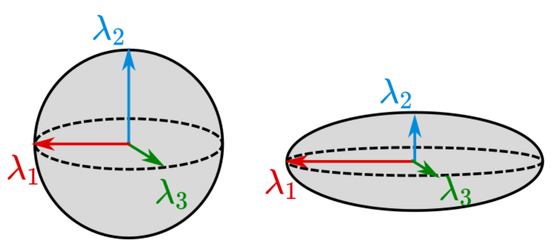

Figure 1: DTI representations.

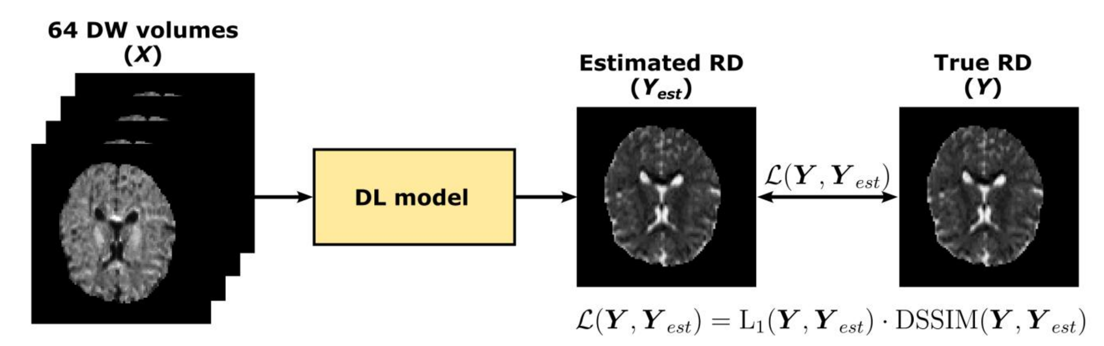

Figure 2: Microstructural parameter estimation using a DL model.

Deep learning: The following settings were used for training:
- model: U-Net architecture (inspired by [3]);
- gradient directions:
  - 64 directions → original data,
  - 32 and 6 directions → simulated from the original data using spherical harmonics;
- training strategy: 100 runs per setup with different weight initializations.

Tract-based spatial statistics (TBSS): The TBSS [4] analysis in FSL [5] involved nonlinear FA alignment to a common space, thresholding at 0.2 to create a mean FA skeleton, and projecting aligned RD data onto it. Voxel-wise cross-subject statistics ($p < 0.05$, 5000 permutations) were performed, with regions of interest defined using the JHU ICBM white matter atlas [6].

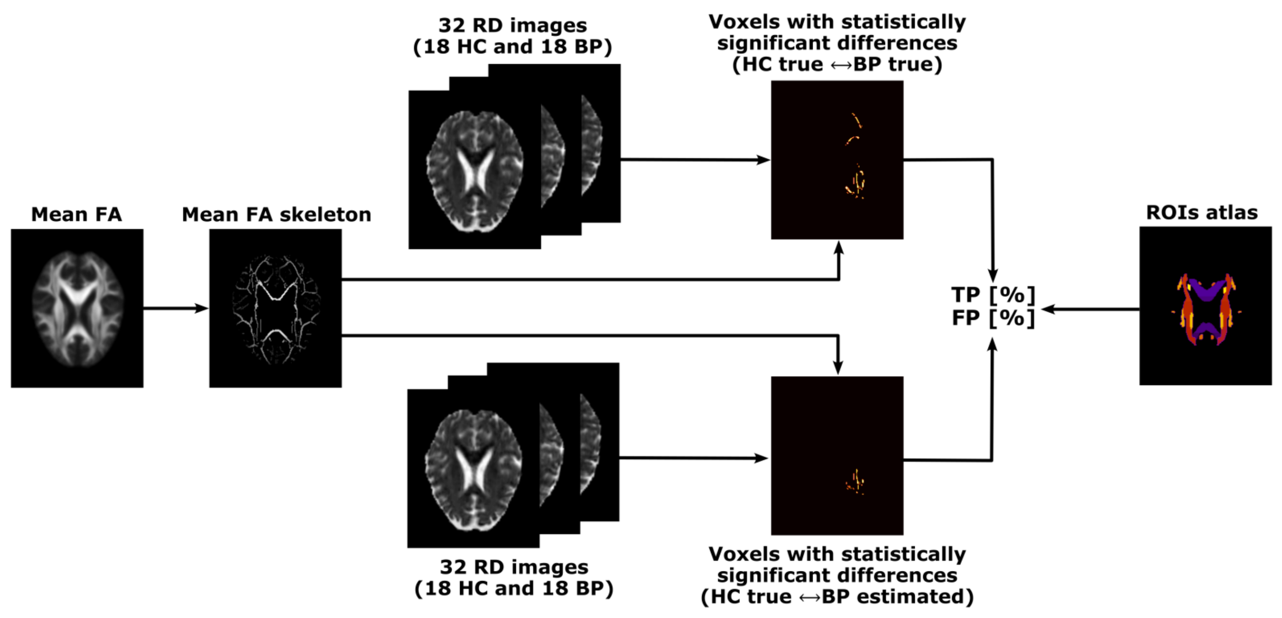

Figure 3: TBSS scheme.

Data: The study used the UCLA dataset [7], including HC and patients with BP.

RESULTS

Table 1: Computation time (36 subjects, 5760 samples).

Runtime environment	CPUs		GPUs	
	Local computer	Supercomputer Ares	Supercomputer Ares	Supercomputer Athena
Single model	~121h	~21h	~38 minutes	~35 minutes
Full analysis (300 models)	~36,300h (50 months)	~168h (7 days)	~5h	~4h 30min

Table 2: Averaged results in the entire white matter on BP data for models trained in three different configurations

		Number of gradient directions		
		6	32	64
Evaluation metrics	MSE [10^{-6}]	108.9 ± 10.5	66.3 ± 10.3	6.3 ± 1.3
	MSSIM	0.9811 ± 0.0013	0.9881 ± 0.0015	0.9994 ± 0.0001
TBSS analysis	TPR [%]	98.54 ± 2.01	98.53 ± 4.70	76.39 ± 30.42
	FPR [%]	342.38 ± 112.81	331.35 ± 123.92	61.62 ± 80.73

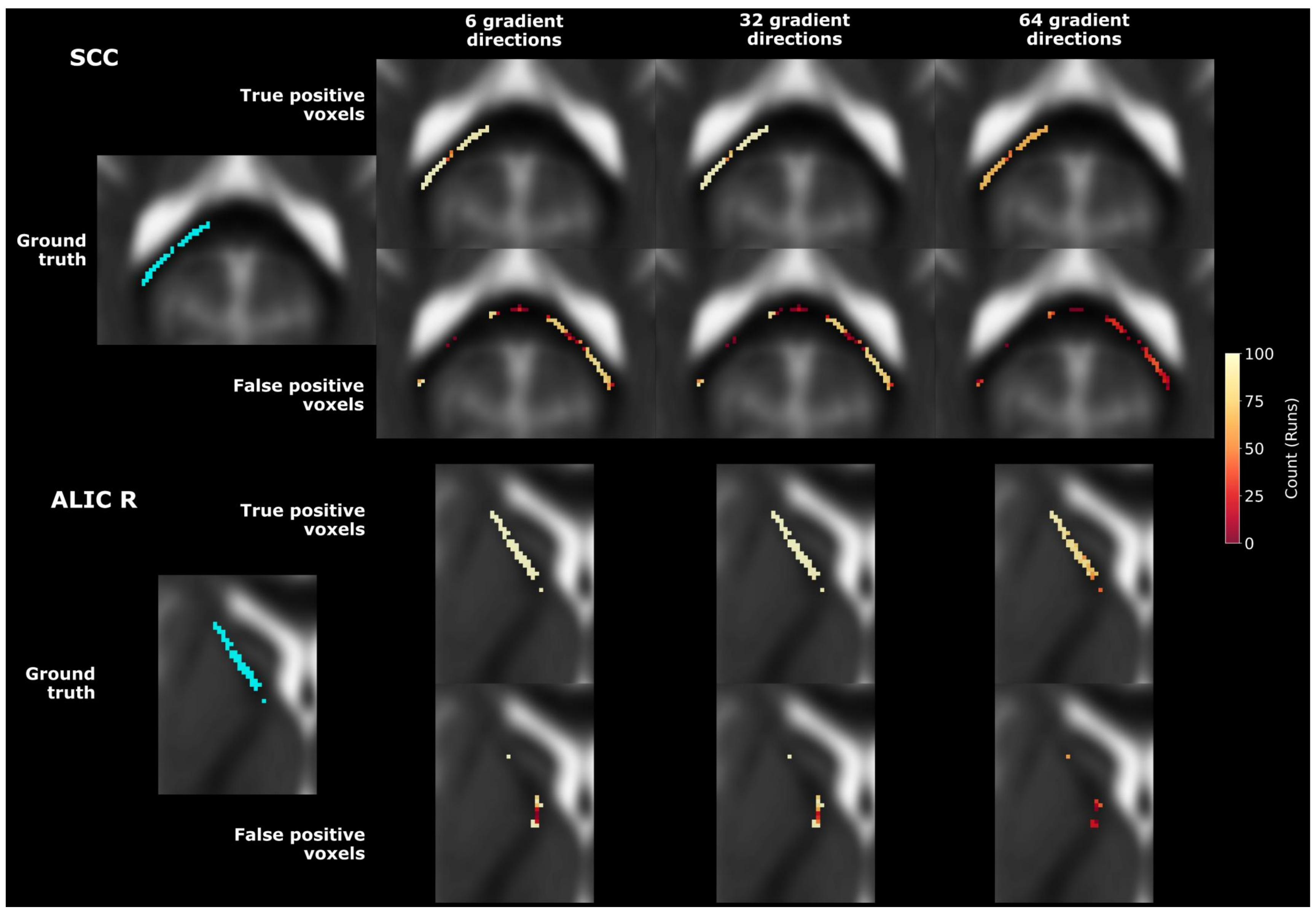

Figure 4: Visualization of true positive and false positive voxels for models trained in three different configurations.

CONCLUSIONS

- DL models achieved high performance based on standard image similarity metrics. However, these metrics did not reflect clinical reliability.
- Variations in model weight initialization had little effect on evaluation metrics, but resulted in significant differences in the TBSS analysis. This underscores the necessity for stability analyses in medical imaging studies.
- Large-scale stability analysis was made feasible through GPU-based parallelization, reducing computation time from years to hours.

REFERENCES

[1] M. L. Phillips et al., The Lancet, 2013.
[2] P. J. Basser et al., Biophysical journal, 1994.
[3] M. Gaviraghi et al., Frontiers in Neuroinformatics, 2022.
[4] S.M. Smith et al., NeuroImage, 2006.
[5] S.M. Smith et al., NeuroImage, 2004.
[6] S. Mori et al., Elsevier, 2005.
[7] K. J. Gorgolewski et al., F1000Research, 2017.

The numerical experiment was possible through computing allocation on the Ares and Athena systems at ACC Cyfronet AGH under the grants PLG/2025/018009 and PLG/2026/019157. Tomasz Pieciak acknowledges the NAWA for grant PPN/BEK/2019/1/00421 and the Ministry of Science and Higher Education under the scholarship for outstanding young scientists (692/STY/13/2018).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857533 and from the International Research Agendas Programme of the Foundation for Polish Science No MAB PLUS/2019/13.

Ministry of Science and Higher Education Republic of Poland

European Funds Smart Growth

Republic of Poland

FNP Foundation for Polish Science

European Union European Regional Development Fund

Funded by the initiative of the Minister of Science and Higher Education "Support for the activities of Centers of Excellence established under the Horizon 2020 program"