A Comparison of Fully-Automated Segmentation Pipelines: QyScore® vs Freesurfer vs FSL vs ANTs

Luca M. Villa[1], Philippe Tran [1], Ayoub Gueddou [1],Nicolas Guizard[1], and Elizabeth Gordon[1]
[1]Qynapse, 2-10 rue d'Oradour-sur-Glane, 75015 Paris, France

QYNAPSE

BACKGROUND

- The gold-standard for brain segmentation is expert manual segmentation, which is both time-consuming and prone to low inter-rater reliability.
- Significant work refining fully-automated segmentation pipelines has resulted in many now providing fast, reliable and accurate segmentations - though performances can vary.
- QyScore® is a fully-automated CE-Marked and FDA-cleared medical imaging platform, certified for grey and white matter segmentation.
- We compared QyScore®'s performance to other freely available state-of-the-art segmentation pipelines, including FreeSurfer, FSL, and ANTs.

METHODS

- A total of 54 T1-weighted images were used for manual and automatic segmentations of the following brain regions: whole brain grey matter (*n*=30), whole brain white matter (*n*=30), hippocampi (*n*=48), amygdala (*n*=48), brainstem (*n*=49), cerebellum (*n*=49), caudate (*n*=49), putamen (*n*=49), thalamus (*n*=49), globus pallidus (*n*=49), and lateral ventricles (*n*=49) (*Table 1*).
- Automated segmentations were produced by QyScore® v1.13, FreeSurfer v7.4.1, FSL v6.0.6.2, and ANTs v2.5, using default parameters, followed by parameter optimization in the instance of preprocessing or segmentation failures. Consensus manual segmentations, created by three expert neuroradiologists, were used as ground truth.
- The dice similarity coefficient (DSC) between manual gold-standard consensus and automated segmentations were used to compare the accuracy of each automated segmentation pipeline.

RESULTS

- Initial automated preprocessing and segmentation failed and required parameter optimization for 46.3% of subjects when using FreeSurfer, and 11.1% of subjects when using FSL. The preprocessing and segmentation of one subject failed for FreeSurfer and FSL and were not recoverable without manual intervention. All subjects were preprocessed and segmented without error, using initial default parameters, by QyScore® and ANTs.

RESULTS

- QyScore® was the best performing segmentation approach for whole brain grey matter, whole brain white matter, brainstem, amygdala, putamen, and lateral ventricles (QyScore® DSC's: 0.79-0.93; ANTs DSC's: 0.73-0.85; FreeSurfer DSC's: 0.61-0.82; FSL DSC's: 0.75-0.83), p_{bonf}<0.001; *Figure 1*.
- FSL performed the best for hippocampal segmentation (DSC=0.83), followed by QyScore® (DSC=0.82), and FreeSurfer (DSC=0.67), p_{bonf}<0.05. There were no significant differences in cerebellum, caudate, globus pallidus, or thalamus segmentation.

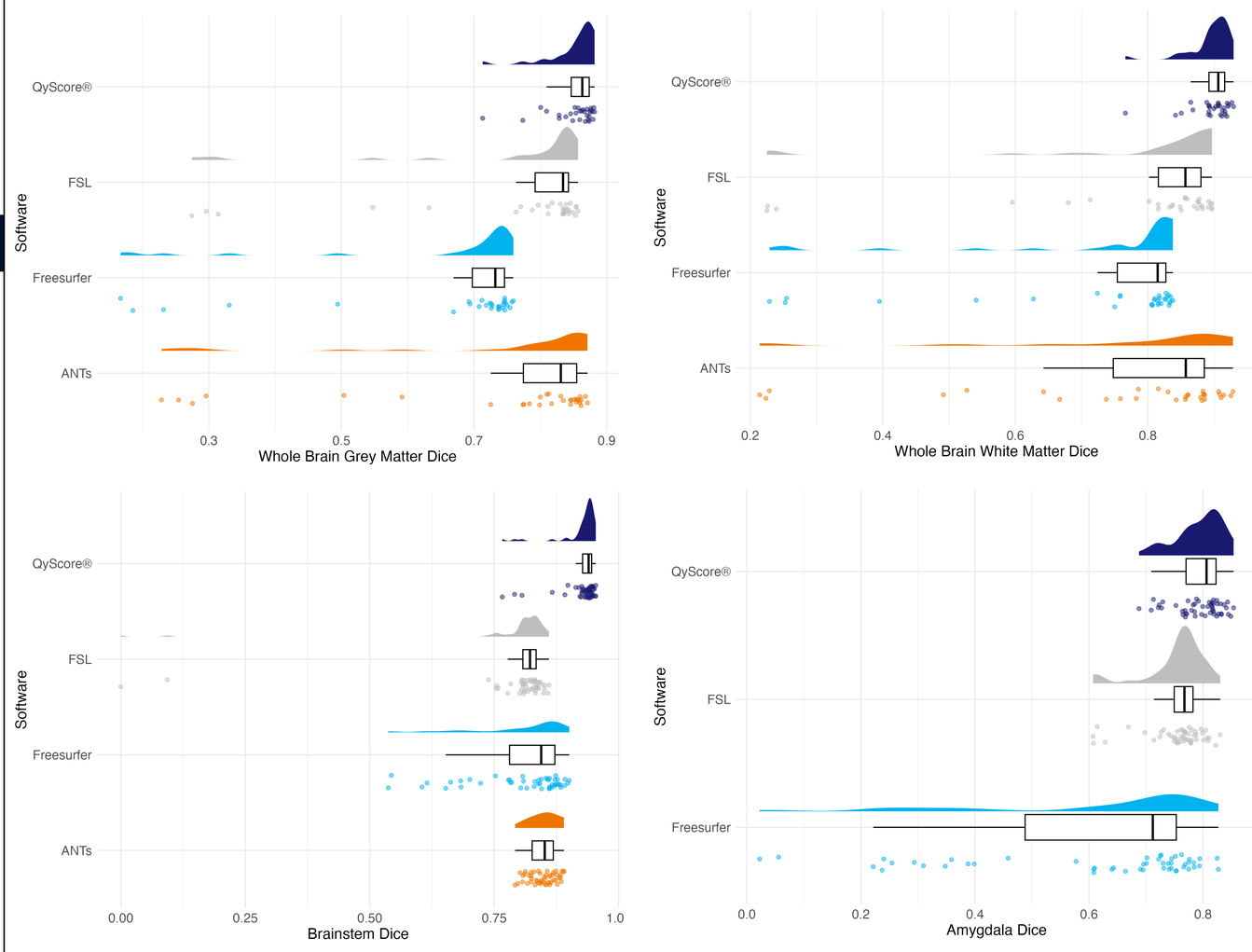

Figure 1. Shows the dice similarity coefficients across brain segmentations for QyScore®, FSL, Freesurfer, and ANTs.

Table 1. Sample clinical and demographic details.

Age (SD), Range	Sex (% Female)	Diagnosis	Field Strength
49.8 (21.5), 18-90	51.9	Healthy control: 77.8% Alzheimer's Disease: 7.4% Multiple Sclerosis: 9.3% Parkinson's Disease: 5.6%	1.5T: 53.7% 3T: 46.3%

CONCLUSIONS

QyScore® showed excellent segmentation accuracy across all brain regions, outperforming the other segmentation pipelines on the majority of regions. Moreover, alongside ANTs, it was the most robust pipeline in processing the wide variety of medical images without error. Its high segmentation accuracy and robust pipeline demonstrate its utility in both healthcare and clinical trial settings, for supporting both clinical diagnosis and monitoring.