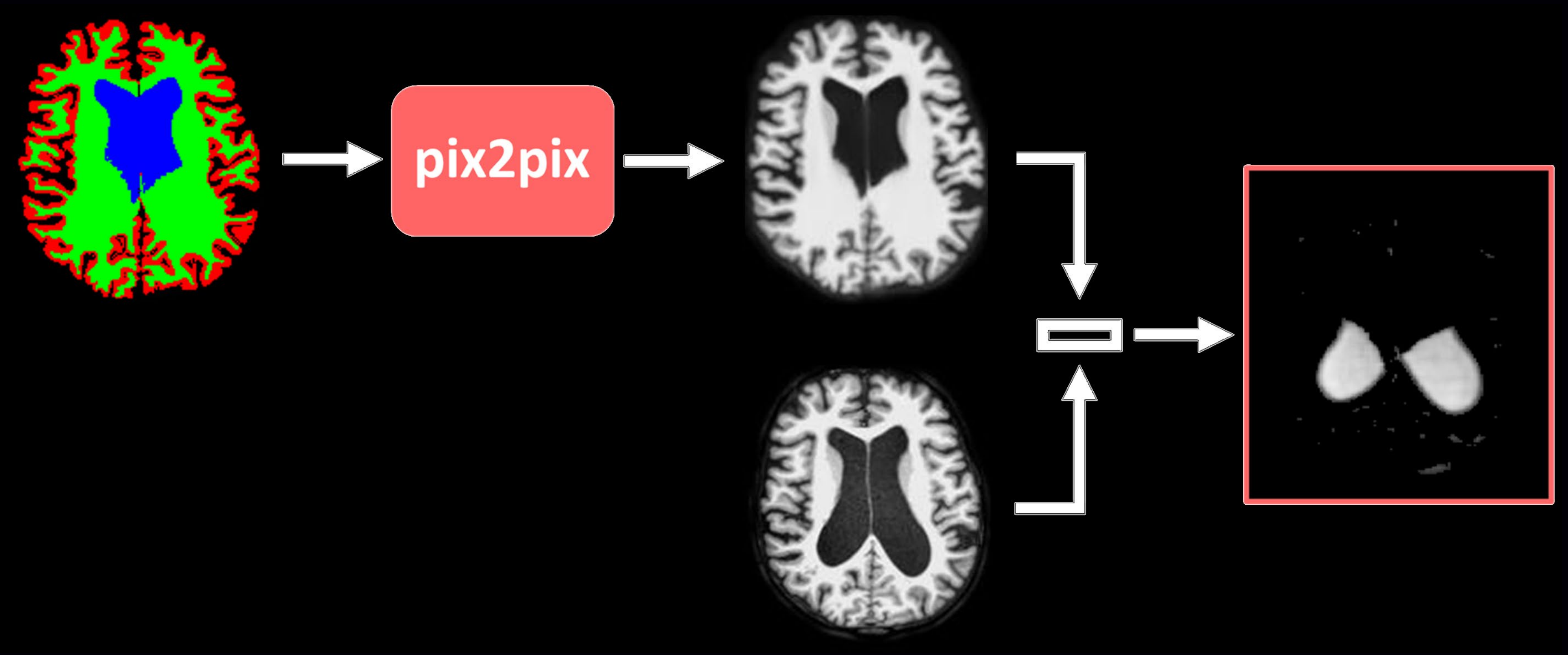
A deep learning-based pipeline for error detection and quality control of brain MRI segmentation results

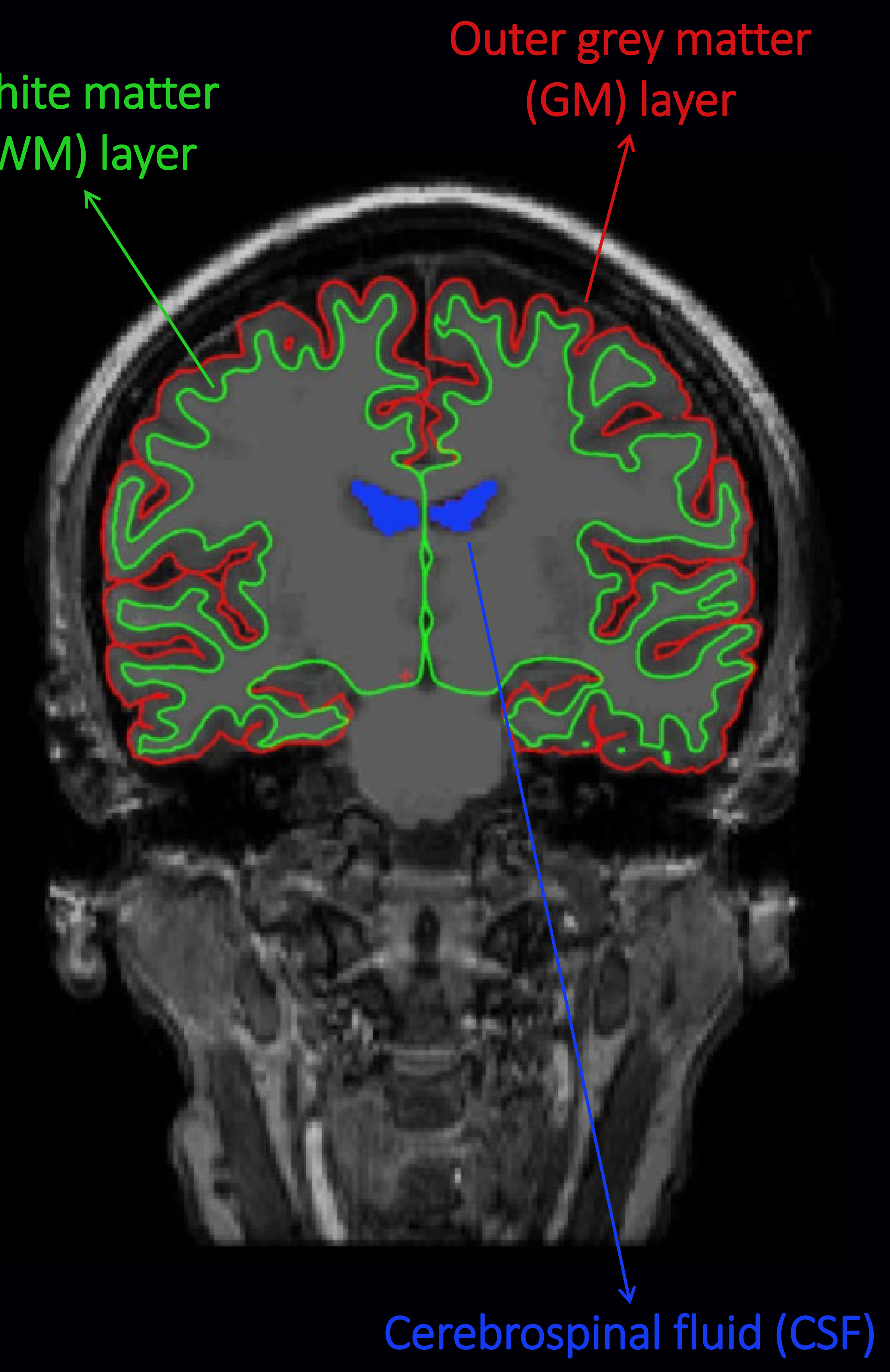


I. Brusini[1,2], D. F. Padilla[2], J. Barroso[3], I. Skoog[4], Ö. Smedby[1], E. Westman[2], C. Wang[1]

[1] KTH Royal Institute of Technology (Stockholm, Sweden), [2] Karolinska Institute (Stockholm, Sweden),
[3] Universidad de La Laguna (La Laguna, Spain), [4] University of Gothenburg (Gothenburg, Sweden)

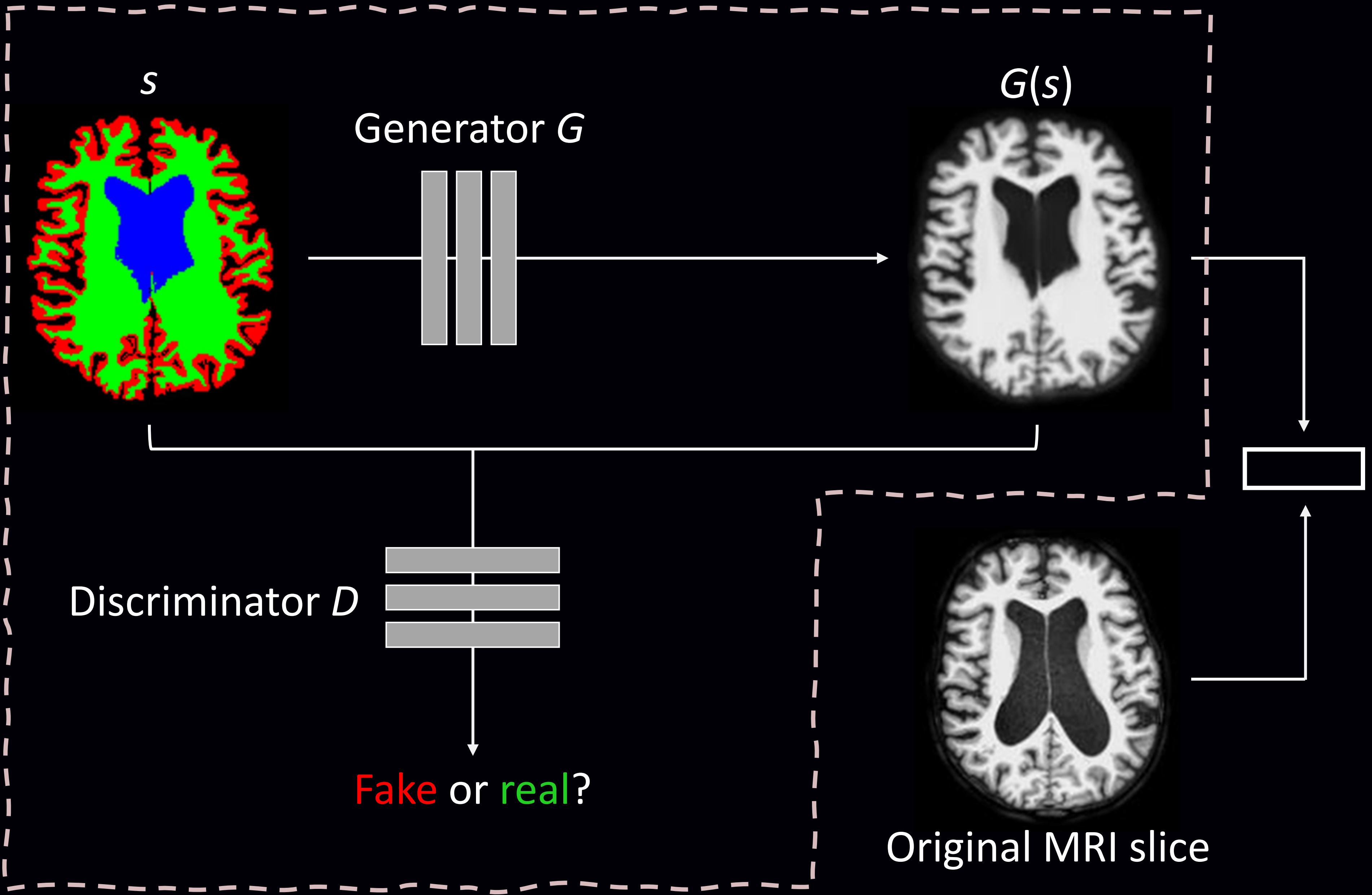
Brain MRI segmentation

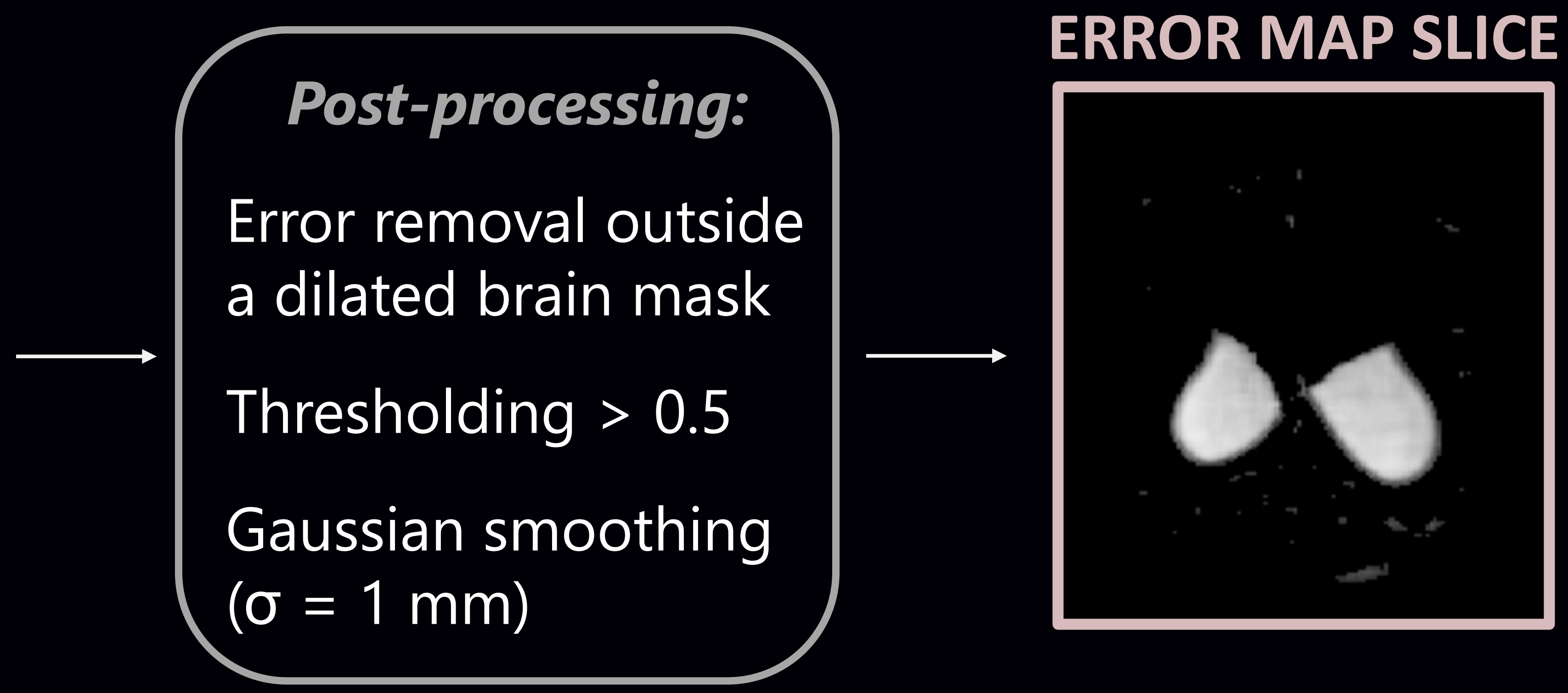
- Useful to study different properties of brain structures (e.g. cortical thickness, ventricular volume, etc.)
- Many tools available for automatic segmentation, but prone to errors
- Quality control (QC) is needed, but it can be very slow and subjective if performed visually



1. Creation of error maps

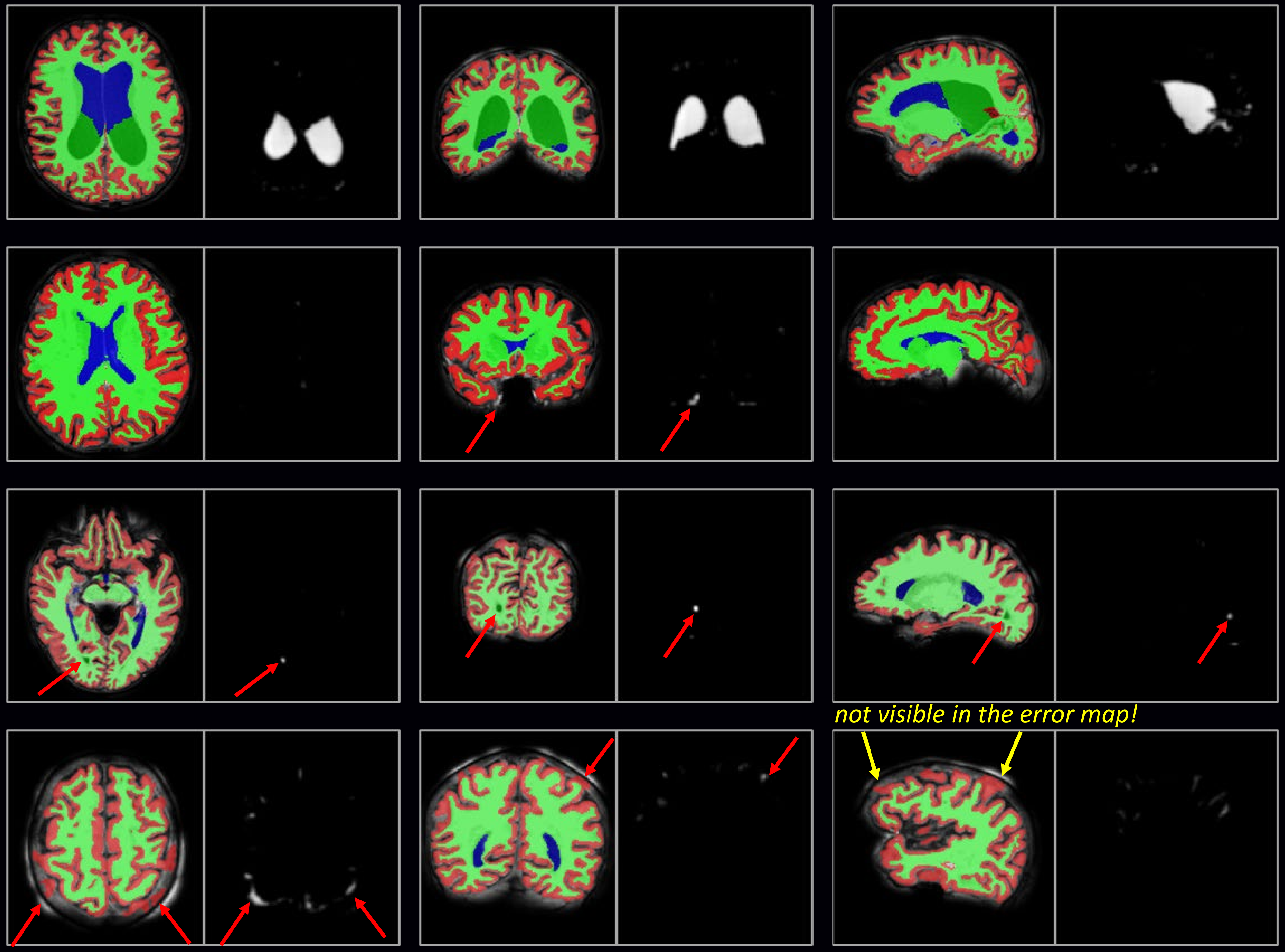
Data from ADNI [1], GENIC [2] and H70 [3] cohorts.
Input segmentations *s* from FreeSurfer [4].





One *pix2pix* model [5] for each view (axial, coronal, sagittal). All trained using random slices from 1600 segmentations that had been visually rated as good.

[1] Jack et al. (2008), *J Magn Reson Imaging*; [2] Ferreira et al. (2015), *Psichotema*; [3] Thorvaldsson et al. (2008), *Neurology;* [4] Fischl et al. (2012), *NeuroImage;*[5] Isola et al. (2017), *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.*



Enlarged ventricles from an Alzheimer's patient had been misclassified and correctly highlighted in the error map.

The error map shows a false positive that is related to subject-specific anatomical features. Other areas of the brain are dark and error-free.

Identification of a small CSF misclassification in a segmentation that had been visually rated as good.

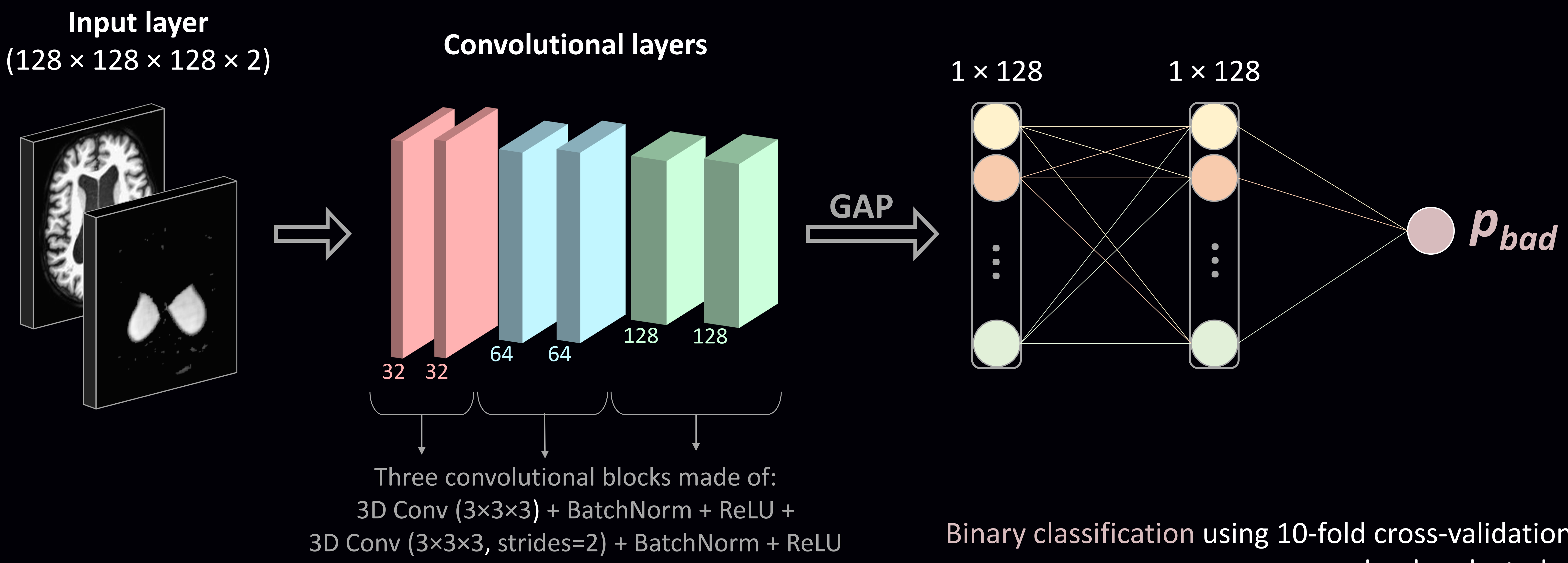
Several cortical overestimations are visible, but the present method highlighted only some of them in the error map.

Promising results for checking segmentation quality, evaluating its reliability and speeding the error correction up.

Limitations: frequent false positives, and some false negatives in correspondence of cortical overestimations.

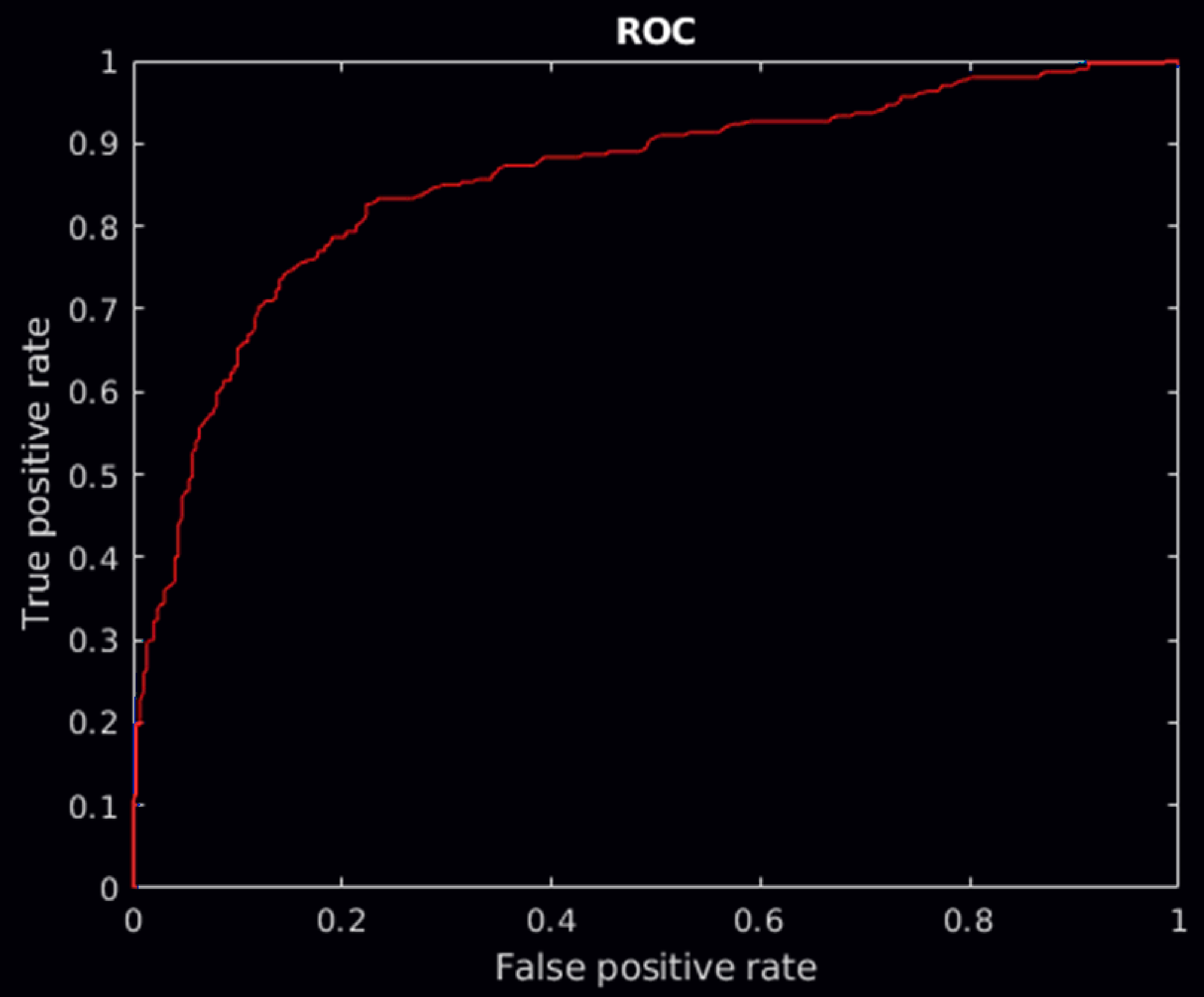
Future work: better post-processing and testing other methods to compare generated and original MRI slices.

2. Classification of segmentation quality



Binary classification using 10-fold cross-validation on 600 randomly selected subjects:
300 with segmentations rated as bad
and 300 rated as good.

Classification performance



AUC = 0.85

High sensitivity (0.96 with a classification threshold of 0.3) is achieved with the proposed classifier, which can be used together with the error maps.

Limitation: high presence of false positives (i.e. segmentations wrongly classified as bad).

However, some errors were missed by the visual raters, so high sensitivity is here preferable to high specificity.