Explainable deep learning based detection of Parkinson's changes in MRI brain scans

Megan Courtman[1], Mark Thurston[2], Lucy McGavin[2], Camille Carroll[1], Lingfen Sun[1], Emmanuel Ifeachor[1], Stephen Mullin[3]

[1]University of Plymouth; [2]University Hospitals Plymouth NHS Trust; [3]Plymouth Institute of Health and Care Research

Background

Candidate neuroprotective treatments for PD are highlighting the **need for early diagnostic tests**. A number of exploratory imaging techniques have suggested that **early pathological brain changes** may be detectable using dedicated experimental MRI sequences.

We explored whether **machine learning** (ML) might be employed to detect such brain changes on routine MRI scans. A subset of ML known as **deep learning** (DL) has shown great promise in diagnostic medical imaging, sometimes matching or even exceeding the diagnostic performance of radiologists.

DL offers the potential of **automated diagnosis** by detecting patterns that might be invisible to the human eye. DL methods have sometimes been criticised for being "black boxes", but newly emerging **explainability methods** are allowing the decisions made by DL models to be better interpreted.

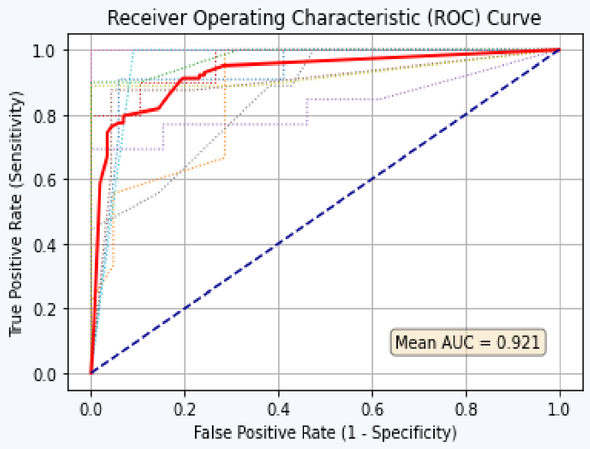
Methods

We trained **convolutional neural networks** to classify T2 axial MRI brain images acquired from the Parkinson's Progression Marker Initiative (PPMI) database.

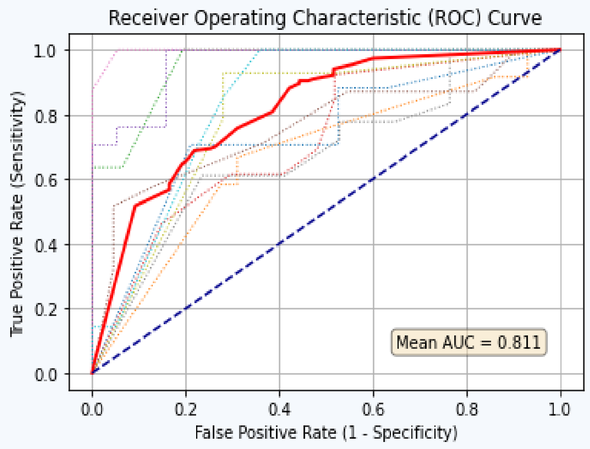
Models were developed for **four PD cohorts stratified by time since diagnosis**: 194 scans acquired more than four years post-diagnosis, 265 acquired two-to-four years post-diagnosis, 241 acquired one-to-two years post-diagnosis, and 282 acquired less than a year post-diagnosis. Each cohort was matched with controls based on age and sex.

Models were assessed using **5-fold cross-validation**. We used **Deep SHapley Additive exPlanations (DeepSHAP)** to calculate and visualise the contribution of individual pixels to the model's prediction.

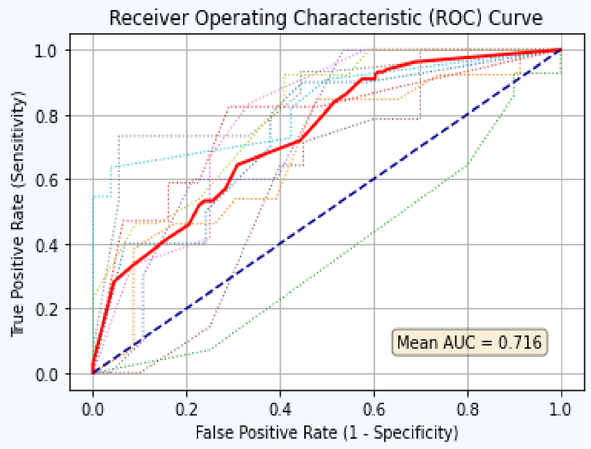
Results



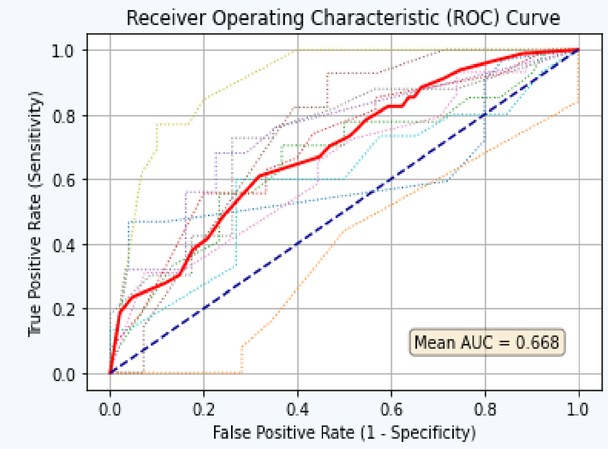
Models trained on scans acquired **over four years** post-diagnosis classified images with **86% accuracy** and 0.92 ROC area-under-the-curve (AUC).



Models trained on scans acquired **two to four years** post-diagnosis classified images with **76% accuracy** and 0.81 ROC AUC.



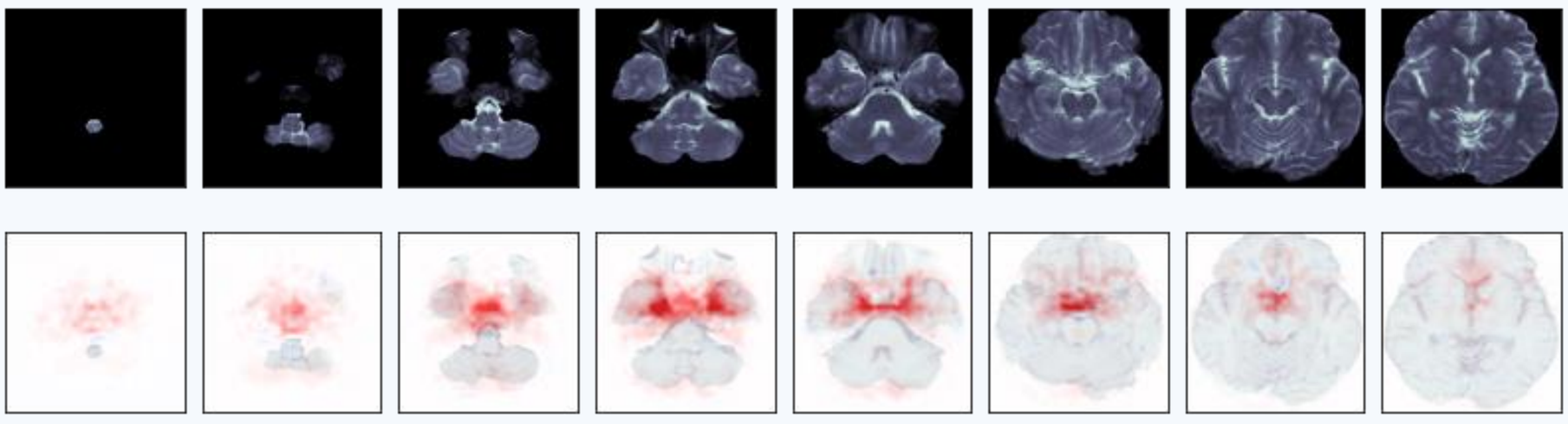
Models trained on scans acquired **one to two years** post-diagnosis classified images with **69% accuracy** and 0.72 ROC AUC.



Models trained on scans acquired **less than a year** post-diagnosis classified images with **64% accuracy** and 0.67 ROC AUC.

The SHAP heatmaps for all cohorts demonstrated **predominant contribution to the classification in the midbrain slices**, as seen in the average heatmap for the models trained on scans acquired over four years post-diagnosis:

Red: pixels that contribute to the model's prediction of PD



Conclusion

The models trained on the later cases of PD exhibited **good diagnostic performance**. The **decreasing performance** for earlier stages of PD suggests that **progressive changes** have been detected. The use of explainable AI highlighted **regions of interest consistent with the known neuropathology** of PD, providing a focus for future work.

We aim to **validate** the results of this pilot study in a large dataset comprised of **routine brain imaging** from National Health Service patients in the South West of England.

UNIVERSITY OF PLYMOUTH

Engineering and Physical Sciences Research Council

NIHR | National Institute for Health Research

NHS University Hospitals Plymouth NHS Trust