Abstract

Machine learning offers great potential for automated prediction of post-stroke symptoms and their response to rehabilitation. Major challenges for this endeavour include the very high dimensionality of neuroimaging data, the relatively small size of the datasets available for learning and interpreting the predictive features, as well as, how to effectively combine neuroimaging and tabular data (e.g. demographic information and clinical characteristics). This paper evaluates several solutions based on two strategies. The first is to use 2D images that summarise MRI scans. The second is to select key features that improve classification accuracy. Additionally, we introduce the novel approach of training a convolutional neural network (CNN) on images that combine regions-of-interests (ROIs) extracted from MRIs, with symbolic representations of tabular data.

We evaluate a series of CNN architectures (both 2D and a 3D) that are trained on different representations of MRI and tabular data, to predict whether a composite measure of post-stroke spoken picture description ability is in the aphasic or non-aphasic range. MRI and tabular data were acquired from 758 English speaking stroke survivors who participated in the PLORAS study. Each participant was assigned to one of five different groups that were matched for initial severity of symptoms, recovery time, left lesion size and the months or years post-stroke that spoken description scores were collected. Training and validation were carried out on the first four groups. The fifth (lock-box/test set) group was used to test how well model accuracy generalises to new (unseen) data.

The classification accuracy for a baseline logistic regression was 0.678 based on lesion size alone, rising to 0.757 and 0.813 when initial symptom severity and recovery time were successively added. The highest classification accuracy (0.854), area under the curve (0.899) and F1 score (0.901) were observed when 8 regions of interest were extracted from each MRI scan and combined with lesion size, initial severity and recovery time in a 2D Residual Neural Network (ResNet). This was also the best model when data were limited to the 286 participants with moderate or severe initial aphasia (with area under curve = 0.865), a group that would be considered more difficult to classify.

Our findings demonstrate how imaging and tabular data can be combined to achieve high post-stroke classification accuracy, even when the dataset is small in machine

learning terms. We conclude by proposing how the current models could be improved to achieve even higher levels of accuracy using images from hospital scanners.