875 Using Stress Testing to Identify Vulnerabilities in Artificial Intelligence Models for the Identification of Culprit Carotid Lesions in Cerebrovascular Events

E. Le[1], J. Tarkin[1], N. Evans[1], M. Chowdhury[2,1], J. Rudd[1]
[1]*Department of Medicine, Addenbrooke's Hospital, Cambridge, United Kingdom,* [2]*Division of Vascular and Endovascular Surgery, Addenbrooke's Hospital, Cambridge, United Kingdom*

Introduction: Carotid atherosclerosis is a major risk factor for ischaemic stroke, a leading cause of death. Carotid CT angiography (CTA) is commonly performed following a stroke or transient ischaemic attack (TIA) to help guide patient management in secondary prevention of stroke. Deep learning algorithms can help extract greater information from scans.

Method: The dataset comprised CTA scans from 40 culprit and 40 non-culprit carotid arteries of patients with recent stroke/TIA, and 40 carotid arteries of asymptomatic patients without previous stroke/TIA. A 3D convolutional neural network was trained to classify carotid artery type. Each input comprised 14 axial CTA carotid patches (centred around the carotid artery) concatenated together to form a 3D volume (capturing ~3cm of artery). 75% of the dataset was used for training and 25% for internal validation. Following training, computer vision operations were applied to input images to assess their impact on the model's classification decisions.

Results: The model achieved 100% accuracy on the training set and 67% on the internal validation set. However, after subjecting input images to image operations, vulnerabilities in the deep learning model were revealed, even when using input images from the training set. For example, using a Gaussian blur filter with sigma 1.0 was sufficient to change classification decisions, as was horizontally flipping the image.

Conclusions: Deep learning has exceptional capabilities for learning, however the risk with such high-capacity models is failure to learn relevant features from the data. Stress testing provides a viable method to further evaluate deep learning models before clinical deployment.