NEAR FOCUS NARROW AND IMAGING DRIVEN ARTIFICIAL INTELLIGENCE FOR THE DIAGNOSIS OF GASTRO-OESOPHAGEAL REFUX DISEASE

1Shraddha Gulati*, 1Julius Bernth, 1Junkai Liao, 1Dmytro Poliyivets, 1Sukhdev Chatu, 1Andrew Emmanuel, 1Amyn Haji, 1Hongbin Lu, 2Bu’Hussein Hayee. 1King’s College Hospital, London, UK; 2King’s College London, London, UK

Introduction Gastro-oesophageal reflux disease (GORD) is a common condition carrying an arduous process for diagnosis. Symptom questionnaires and proton-pump inhibitor challenge lack reliability and white light endoscopy (WLE) is often normal. Objective measurement of oesophageal acid exposure time (AET) require invasive testing. Changes in intrapapillary capillary loops (IPCLs) identified using narrow-band imaging (NBI) have been proposed as a marker for reflux. We evaluated a near focus (NF-NBI) driven artificial intelligence (AI) model for the diagnosis of GORD.

Methods Patients with symptoms of GORD (recorded using the Reflux Disease Questionnaire (RDQ)) were prospectively recruited over 10 months. Upper endoscopy recorded multiple NF-NBI images, video and biopsies of the lower oesophagus. If endoscopy using High-Definition WLE was normal, a pH-recording capsule was placed. Patients were defined according to Lyon criteria; Erosive oesophagitis (EO); non-erosive reflux disease (NERD); functional heartburn (FH).

Two forms of AI were developed and evaluated to automate regions of interest (ROI) and detect IPCLs and morphological features using computer vision (CV) and deep convolutional neural network (DCNN) using Resnet50. DCNN was evaluated using training: unseen testing dataset ratios of 50:50 (3872:4280 images) and 75:25 (6484:1668 images). For the purposes of training the AI models, EO and NERD cases were combined as ‘GORD’. A novel combined classifier (CC) of both AI methods was evaluated.

Results 78 consecutive patients were recruited. n=68 (46 Female, 44.41±12.91 years): GORD n=27 (EO n=6, NERD n=21) and FH n=41 were analysed. The mean IPCL per ROI count was greater in GORD vs FH: 33.2 ± 5.19 vs 28.1 ± 5.42 p=0.0003 and was used as the primary diagnostic tool. IPCL morphology for GORD vs FH: length 16.29 vs 16.98, p=0.19; width 7.8 vs 7.8, p=0.98; red 118.8 vs 120.6, p=0.44; green 114.3 vs 118, p=0.004; blue 94.24 vs 97.54 p=0.07.

With CV: mean IPCLs/ROI (threshold 27.6) had sensitivity, specificity, AUC: 88.9, 58.5, 0.76 (p=0.0003) for GORD. With DCNN 50:50 these results were 58%, 86% and 76% respectively. DCNN 75:25 produced 67%, 92%, 83% respectively.

CC improved overall specificity (89.1%) and accuracy (78.1%) but not sensitivity (63%).

Conclusion AI using NF-NBI is a novel method for the diagnosis of GORD. With increased data, improvements in diagnostic accuracy is achieved further improved using a CC. This model has the potential to provide a reliable safe single-test diagnosis of GORD.