## **The first Virtual Chromoendoscopy Artificial Intelligence system to detect endoscopic and histologic healing in Ulcerative Colitis**

**Authors**: Marietta Iacucci, Rosanna Cannatelli, Tommaso Lorenzo Parigi, Andrea Buda, Nunzia La Barile, Olga Maria Nardone, Gian Eugenio Tontini, Alessandro Rimondi, Alina Bazarova, Pradeep Bhandari, Raf Bisschops, Gert De Hertogh, Rocio del Amor, Jose G Ferraz, Martin Goetz, Xianyong Gui, Bu Hayee, Ralf Kiesslich, Mark Lazarev,Valery Naranjo, Remo Panaccione, Adolfo Parra-Blanco, Luca Pastorelli, Timo Rath, Elin S Røyset, Michael Vieth, Vincenzo Villanacci, Davide Zardo, Subrata Ghosh,Enrico Grisan

**Background**

Endoscopic and histologic activity are important therapeutic targets in ulcerative colitis (UC). The Paddington International Virtual ChromoendoScopy ScOre (VCE-PICaSSO)1 demonstrated that enhanced visualisation of subtle mucosal and vascular inflammatory changes correlated strongly with histology. However, without adequate training, the subjective evaluation of white light (WL) and VCE endoscopic scores varies between observers. We aimed to develop an artificial intelligence (AI) system for objective assessment of endoscopic disease activity and predict histology related to both white light and VCE videos.

**Methods**

469 endoscopy videos (48512 frames) from 235 patients representative of all grades of inflammation, from our prospective PICaSSO multicentre study1 were used to develop a convolutional neural network (CNN). 316 videos were divided into training (254) and validation (62) sets. 153 additional videos (78 patients) were used as test cohort. The videos were edited to separate clips with WL and with VCE, and assessed using Ulcerative Colitis Endoscopic Index of Severity (UCEIS) and PICaSSO, respectively. The classification stage of a pre-trained ResNet50 CNN classifier was trained to predict the healing or active inflammation on video frames. One network was trained to predict endoscopic remission (ER) as UCEIS≤1 from WL frames, and a second network was trained to predict PICaSSO≤3 from VCE. Histological remission (HR) was defined as Robarts Histological Index (RHI) ≤3 with no neutrophils in lamina propria or epithelium.

**Results**

In the validation cohort, our system predicted ER (UCEIS ≤1) in WL videos with 82% sensitivity (Se), 94% specificity (Sp) and an area under the ROC curve (AUROC) of 0.92. For the detection of remission in VCE videos (PICaSSO ≤3) Se was 74%, Sp 95%, and AUROC 0.95. In the testing cohort of independent videos, the diagnostic performance for both cut offs of ER remained similar.

Our system also had an excellent diagnostic performance for the prediction of HR in the validation set, with Se, Sp, and Accuracy of 92%, 83%, and 85% respectively, using VCE, and 83%, 87%, and 86% respectively, with WL. In the testing set performance declined modestly while remaining good. Of note, the algorithm’s prediction of histology was similar with VCE and WL videos.

**Table 1 Algorithm performance in prediction of endoscopic healing according to PICaSSO score with VCE, and to UCEIS with WL.**

|  |  |  |
| --- | --- | --- |
| **Validation Cohort** | **Endoscopic remission (Picasso 0-3)** | **Endoscopic remission (UCEIS 0-1)** |
|  | 62 VCE videos | 58 WL videos |
| Sensitivity | 0.74 (0.53 - 0.93) | 0.82 (0.67 - 0.98) |
| Specificity | 0.95 (0.89 - 1.02) | 0.94 (0.87 - 1.02) |
| PPV | 0.88 (0.71 - 1.03) | 0.90 (0.78 - 1.03) |
| NPV | 0.89 (0.80 - 0.98) | 0.89 (0.79 - 0.99) |
| Accuracy | 0.89 (0.81 - 0.97) | 0.89 (0.81 - 0.97) |
| Cohen’s Kappa | 0.72 (0.53 - 0.92) | 0.78 (0.61 - 0.95) |
|  |  |  |
| **Testing Cohort** |  |  |
|  | 153 VCE videos | 152 WL videos |
| Sensibility | 0.83 (0.62 - 1.04) | 0.59 (0.35 - 0.82) |
| Specificity | 0.90 (0.85 - 0.96) | 0.87 (0.81 - 0.92) |
| PPV | 0.54 (0.35 - 0.73) | 0.36 (0.18 - 0.53) |
| NPV | 0.87 (0.82 - 0.93) | 0.94 (0.90 - 0.98) |
| Accuracy | 0.82 (0.76 - 0.88) | 0.84 (0.77 - 0.91) |
| Cohen’s Kappa | 0.37 (0.15 - 0.57) | 0.34 (0.12 - 0.59) |

**Table 2 Performance in prediction of histologic healing (Robarts Index) with VCE and WL (validation cohort)**

|  |  |
| --- | --- |
| **Virtual chromoendoscopy** | **Histological remission (RHI 0-3)** |
| **Validation cohort** |  |
|  | 61 videos |
| Specificity | 0.83 (0.72 - 0.94) |
| Sensitivity | 0.92 (0.79 - 1.06) |
| PPV | 0.62 (0.41 - 0.83) |
| NPV | 0.97 (0.93 - 1.02) |
| Accuracy | 0.85 (0.76 – 0.94) |
| Cohen’s Kappa | 0.65 (0.43 - 0.86) |
| **White light** |  |
|  | 59 videos |
| Specificity | 0.87 (0.78 - 0.97) |
| Sensitivity | 0.83 (0.62 - 1.04) |
| PPV | 0.63 (0.39 - 0.86) |
| NPV | 0.95 (0.89 - 1.02) |
| Accuracy | 0.86 (0.77 - 0.95) |
| Cohen’s Kappa | 0.63 (0.39 - 0.87) |

**Figure 1 Artificial intelligence detection of inflammation**

**A picture containing plastic

Description automatically generated**