Introduction Manual review and annotation of a capsule endoscopy (CE) video requires a considerable amount of reviewing time. The diagnostic accuracy of this process may decrease over time due to reviewer tiredness. Recent studies showed an average detection rate – for the clinically significant findings – as low as 40%.1 We present a generic computational framework for automatic detection of abnormalities in CE videos.

Methods A CE video (MirCam®, Intromedic Co Ltd, Seoul, Korea), depicting inflammatory changes (aphthae, mucosal breaks, ulcers, erythema) was reviewed and manually annotated by experienced CE reviewer. A total of 1984 frames, containing any pathology, were thumbnailed. The proposed framework considers video frames as members of a vector space represented by their colour information. An unsupervised data reduction algorithm,2 which does not require any prior knowledge about the data, was then applied on each segment. This algorithm clusters together frames that exhibit similar characteristics e.g., colour distributions. Its output is a subset of video frames extracted from each cluster by applying a threshold to the clustering result. The extracted frames are characteristic of the particular video segment and as a result representative of possible lesions.

Results The evaluation of the proposed framework aimed to determine its accuracy, in terms of the ratio of the neighbourhoods represented by at least one frame in the system’s output and the neighbourhoods that were manually annotated as suspicious for containing lesions. The parameters considered include clustering from 2 to 6 clusters and thresholds varying from 0.004 to 0.6. The obtained accuracy ranged between 76% to 98% depending on the desired sensitivity level of the algorithm, controlled by the threshold. Furthermore, the automatic selection of the representative CE video segments performed by the proposed approach, the number of video frames to be thoroughly examined can be reduced from 30% to 60% of the original video, depending on the clustering and threshold settings.

Conclusion The application of the proposed framework to the evaluation of CE videos may reduce the rate of false negative evaluations by attracting the attention of the reviewer to automatically identified video segments (or single frames) of interest which are likely to contain lesions.