OPEN ACCESS

Vessel and tissue recognition during third-space endoscopy using a deep learning algorithm

Alanna Ebigbo ,[1] Robert Mendel,[2] Markus W Scheppach ,[1] Andreas Probst,[1] Neal Shahidi ,[3] Friederike Prinz,[1] Carola Fleischmann,[1] Christoph Römmele ,[1] Stefan Karl Goelder,[4] Georg Braun,[1] David Rauber,[2] Tobias Rueckert,[2] Luis A de Souza Jr,[5] Joao Papa,[6] Michael Byrne,[7] Christoph Palm,[2] Helmut Messmann[1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-326470).

For numbered affiliations see end of article.

Correspondence to
Dr Alanna Ebigbo, Department of Gastroenterology, Universitätsklinikum Augsburg, Augsburg 86156, Bayern, Germany;
alanna.ebigbo@uk-augsburg.de

AE and RM contributed equally.

Received 1 November 2021
Accepted 29 June 2022

MESSAGE

In this study, we aimed to develop an artificial intelligence clinical decision support solution to mitigate operator-dependent limitations during complex endoscopic procedures such as endoscopic submucosal dissection and peroral endoscopic myotomy, for example, bleeding and perforation. A DeepLabv3-based model was trained to delineate vessels, tissue structures and instruments on endoscopic still images from such procedures. The mean cross-validated Intersection over Union and Dice Score were 63% and 76%, respectively. Applied to standardised video clips from third-space endoscopic procedures, the algorithm showed a mean vessel detection rate of 85% with a false-positive rate of 0.75/min. These performance statistics suggest a potential clinical benefit for procedure safety, time and also training.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Recently, artificial intelligence (AI) tools have been developed for clinical decision support in diagnostic endoscopy, but so far, no algorithm has been introduced for therapeutic interventions.

WHAT THIS STUDY ADDS

⇒ Considering the elevated risk of bleeding and perforation during endoscopic submucosal dissection and peroral endoscopic myotomy, there is an apparent need for innovation and research into AI guidance in order to minimise operator-dependent complications. In this study, we developed a deep learning algorithm for the real-time detection and delineation of relevant structures during third-space endoscopy.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This new technology shows great promise for achieving higher procedure safety and speed. Future research may further expand the scope of AI applications in GI endoscopy.

IN MORE DETAIL

Endoscopic submucosal dissection (ESD) is an established organ-sparing curative endoscopic resection technique for premalignant and superficially invasive neoplasms of the GI tract.[1 2] However, ESD and peroral endoscopic myotomy (POEM) are complex procedures with an elevated risk of operator-dependent adverse events, specifically intraprocedural bleeding and perforation. This is due to inadvertent transection through submucosal vessels or into the muscularis propria, as visualisation and cutting trajectory within the expanding resection defect is not always apparent.[3 4] An effective mitigating strategy for intraprocedural adverse events has yet to be developed.

Artificial intelligence clinical decision support solution (AI-CDSS) has rapidly proliferated throughout diagnostic endoscopy.[5–7] We therefore sought to develop a novel AI-CDSS for real-time intraprocedural detection and delineation of vessels, tissue structures and instruments during ESD and POEM.[8]

Sixteen full-length videos of 12 ESD and 4 POEM procedures using Olympus EVIS X1 series endoscopes (Olympus, Tokyo, Japan) were extracted from the Augsburg University Hospital database. A total of 2012 still images from these videos were annotated by minimally invasive tissue resection experts (ESD experience ≥500 procedures) using the computer vision annotation tool for the categories electrosurgical knife, endoscopic instrument, submucosal layer, muscle layer and blood vessel. A DeepLabv3+ neural network architecture with KSAC[9] and a 101-layer ResNeSt backbone[10] (online supplemental methods) was trained with these data. The performance of the algorithm was measured in an internal fivefold cross validation, as well as a test on 453 annotated images from 11 separate videos using the parameters Intersection over Union (IoU), Dice Score and pixel accuracy (online supplemental methods). The IoU and Dice Score measure the percentual overlap between the algorithm's delineation and the gold standard. The pixel accuracy measures the percentage of true pixel predictions per image and over all classes. The validation metrics were calculated by accumulating the per-fold outputs. The cross validation was done without hyperparameter tuning.

Three further full-length videos (1× POEM, 1× rectal ESD and 1× oesophageal ESD) were used for an evaluation of the algorithm on video. Thirty-one clips with 52 predefined vessels (online supplemental methods) were evaluated frame by frame with artificial intelligence (AI) overlay for true and

To cite: Ebigbo A, Mendel R, Scheppach MW, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2021-326470

Table 1 Performance results of the AI-CDSS in the internal cross validation and the test data set: IoU and Dice Score for all categories as well as their means across all categories, pixel accuracy for complete frames and 95% CI in brackets

Internal cross validation							
	Vessel detection	Tissue differentiation			Instrument detection		
	Vessel	Submucosa	Muscularis	Background	Instrument	Knife	Mean
Dice Score	55.15 (54.10 to 56.18)	75.51 (74.88 to 76.12)	70.64 (69.32 to 71.88)	86.49 (85.99 to 86.99)	88.69 (87.57 to 89.83)	80.60 (79.61 to 81.49)	76.18 (75.73 to 76.57)
IoU	38.07 (37.08 to 39.07)	60.65 (59.85 to 61.44)	54.60 (53.05 to 56.10)	76.19 (75.43 to 76.98)	79.68 (77.89 to 81.54)	67.51 (66.13 to 68.77)	62.78 (62.18 to 63.31)
Pixel accuracy	80.99 (80.52 to 81.47)						
Test							
Dice Score	62.77 (60.08 to 65.12)	80.71 (79.50 to 81.82)	72.48 (69.40 to 74.99)	91.39 (90.45 to 92.10)	89.69 (87.09 to 91.96)	83.50 (82.06 to 84.87)	80.09 (79.14 to 80.92)
IoU	45.74 (42.94 to 48.28)	67.65 (65.97 to 69.24)	56.84 (53.14 to 59.99)	84.14 (82.56 to 85.36)	81.30 (77.14 to 85.11)	71.67 (69.58 to 73.72)	67.89 (66.61 to 69.04)
Pixel accuracy	86.89 (85.86 to 87.70)						

AI-CDSS, artificial intelligence clinical decision support solution.

false vessel detection, and a vessel detection rate (VDR) was determined.

The cross-validated mean IoU, mean Dice Score and pixel accuracy were 63%, 76% and 81%, respectively. On the test set, the AI-CDSS achieved scores of 68%, 80% and 87% for the same parameters. The individual per class values and 95% CIs are shown in table 1. Examples of the original frames, expert annotations and AI segmentations are shown in figure 1.

The mean VDR was 85%. The VDR for rectal ESD, oesophageal ESD and POEM were 70%, 95% and 92%, respectively. The mean false-positive rate was 0.75 /min. The algorithm spotted seven out of nine vessels, which caused intraprocedural bleeding. It also recognised the two vessels which required specific haemostasis by haemostatic forceps for major bleeding.

To demonstrate the performance of the AI-CDSS without computing quantitative performance measures, we show an example of an internal POEM procedure with AI overlay. For visualisation of the experiment, we show six video clips, which were used for the evaluation of VDR in the same video (2× POEM, 2× rectal ESD and 2× oesophageal ESD; online supplemental video 1). For a test in robustness, the algorithm was also applied to a randomly selected highly compressed YouTube video of a gastric per-oral endoscopic myotomy procedure (ENDOCLUNORD 2020, https://www.youtube.com/watch?v=VKF-HWOzYDGM; online supplemental video 2). The individual output is the result of an exponential moving average between the current and past predictions which smoothes the predictions and is a simple way to include temporal information.

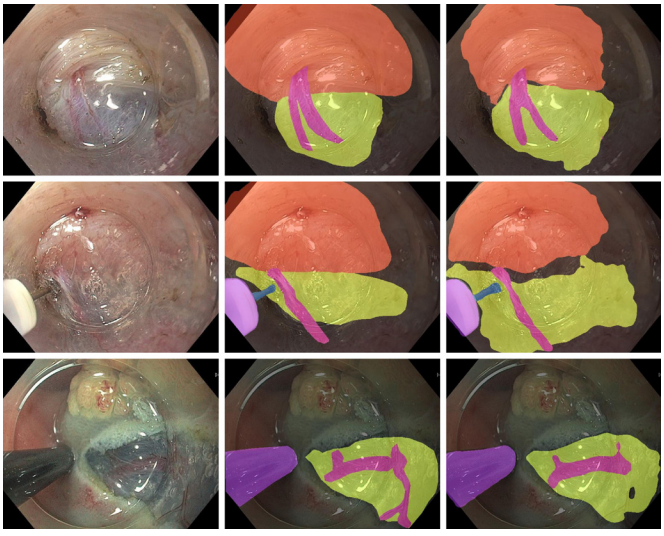

Figure 1 Examples of original images (left column) with corresponding expert annotations (middle column) and AI segmentations (right column). The muscle layer, submucosa, vessels and knife are segmented with a coloured overlay.

COMMENTS

This preliminary study aims at investigating the potential role of AI during therapeutic endoscopic procedures such as ESD or POEM. The algorithm delineated tissue structures, vessels and instruments in frames taken from endoscopic videos with a high overlap to the gold standard provided by expert endoscopists. Analogous technology[11] has been demonstrated for application in laparoscopic cholecystectomy to differentiate between safe and dangerous zones of dissection with a mean IoU of 53% and 71%, respectively.

On video clips with standardised and predefined vessels, the algorithm showed a VDR of 85%. The lower performance of 70% in rectal ESD compared with excellent detection of over 90% in oesophageal ESD and POEM might be explainable by poorer visualisation of the structures and more intraprocedural bleeding, which is in agreement with clinical experience.

Numerous preclinical and clinical studies on AI in GI endoscopy have been published, but until now, the application of AI has been limited largely to diagnostic procedures such as the detection of polyps or the characterisation of unclear lesions. In abdominal surgery, AI has been applied with promising results for various tasks, including the detection of surgical instruments, image guidance, navigation and skill assessment ('smart surgery').[12] The results of this study suggest that AI may have the potential to optimise complex endoscopic procedures such as ESD or POEM in analogy to the mentioned research ('smart ESD'). By highlighting submucosal vessels and other tissue structures, such as the submucosal cutting plane, therapeutic procedures could become faster and burdened with fewer adverse events such as intraprocedural or postprocedural bleeding and

perforation. In the future, AI assistance may have the potential to accelerate the learning curve of trainees in endoscopy.

The major limitation of this study is the small number of videos used for training and validation; however, every video contained a complete therapeutic ESD procedure with a full range of procedural situations. The study is further limited by the fact that the algorithm was not yet tested in a real-life setting. However, the AI model was tested on externally generated video sequences and was able to recognise submucosal vessels and the cutting plane. Furthermore, surrogate parameters such as the detection of vessels, which bled later during the procedures, give rise to the conclusion that these complications might have been preventable by the application of the AI-CDSS. This is a first preclinical report on a novel technology; further research is needed to evaluate a potential clinical benefit of this AI-CDSS in detail.

Author affiliations
[1]Department of Gastroenterology, Universitätsklinikum Augsburg, Augsburg, Germany
[2]Regensburg Medical Image Computing (ReMIC), Ostbayerische Technische Hochschule Regensburg, Regensburg, Germany
[3]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada
[4]Department of Gastroenterology, Ostalb-Klinikum Aalen, Aalen, Germany
[5]Department of Computing, Federal University of São Carlos, São Carlos, Brazil
[6]Department of Computing, São Paulo State University, Botucatu, Brazil
[7]Vancouver General Hospital, The University of British Columbia, Vancouver, British Columbia, Canada

Twitter Joao Papa @papa_joaopaulo and Christoph Palm @ReMIC_OTH

Contributors AE and MWS: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript and critical revision of the manuscript. RM: study concept and design, software implementation, analysis and interpretation of data, drafting of the manuscript and critical revision of the manuscript. AP: study concept and design, acquisition of data and critical revision of the manuscript. NS: analysis and interpretation of data, drafting of the manuscript and critical revision of the manuscript. FP, CF, CR, SKG and GB: acquisition of data and critical revision of the manuscript. DR and TR: software implementation and experimental evaluation and critical revision of the manuscript. LAdS: statistical analysis and critical revision of the manuscript. JP: Statistical analysis, critical revision of the manuscript and study supervision. MB: study concept and design, analysis and interpretation of data, drafting of the manuscript and critical revision of the manuscript. CP: study concept and design, analysis and interpretation of data, statistical analysis, critical revision of the manuscript, administrative and technical support and study supervision. HM: study concept and design, acquisition of data, critical revision of the manuscript, administrative and technical support, and study supervision. AE: guarantor. All authors: had access to the study data and reviewed and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests NS: speaker honorarium, Boston Scientific and Pharmascience. MB: CEO and founder, Satisfai Health. HM: consulting fees, Olympus.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable. Ethics approval was obtained from the ethics committee of Ludwigs-Maximilians-Universität, Munich (project number 21–1216).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Alanna Ebigbo http://orcid.org/0000-0001-7765-035X
Markus W Scheppach http://orcid.org/0000-0002-2255-7032
Neal Shahidi http://orcid.org/0000-0002-4536-0515
Christoph Römmele http://orcid.org/0000-0002-8745-8510

REFERENCES

1 Bourke MJ, Neuhaus H, Bergman JJ. Endoscopic submucosal dissection: indications and application in Western endoscopy practice. *Gastroenterology* 2018;154:1887–900.
2 Shahidi N, Rex DK, Kaltenbach T, *et al*. Use of endoscopic impression, artificial intelligence, and pathologist interpretation to resolve discrepancies between endoscopy and pathology analyses of diminutive colorectal polyps. *Gastroenterology* 2020;158:783–5.
3 Draganov PV, Aihara H, Karasik MS, *et al*. Endoscopic submucosal dissection in North America: a large prospective multicenter study. *Gastroenterology* 2021;160:2317–27.
4 Fleischmann C, Probst A, Ebigbo A, *et al*. Endoscopic submucosal dissection in Europe: results of 1000 neoplastic lesions from the German endoscopic submucosal dissection registry. *Gastroenterology* 2021;161:1168–78.
5 Byrne MF, Chapados N, Soudan F, *et al*. Real-Time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94–100.
6 Ebigbo A, Mendel R, Probst A, *et al*. Computer-Aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut* 2019;68:1143–5.
7 Shahidi N, Bourke MJ. How to manage the large Nonpedunculated colorectal polyp. *Gastroenterology* 2021;160:2239–43.
8 Khashab MA, Benias PC, Swanstrom LL. Endoscopic myotomy for foregut motility disorders. *Gastroenterology* 2018;154:1901–10.
9 Huang Y, Wang Q, Jia W. See More Than Once - Kernel-Sharing Atrous Convolution for Semantic Segmentation. *arXiv* 2019;190809443.
10 Zhang H, Wu C, Zhang Z. ResNeSt: Split-Attention networks. *arXiv* 2020;200408955.
11 Madani A, Namazi B, Altieri MS, *et al*. Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. *Ann Surg* 2020. doi:10.1097/SLA.0000000000004594. [Epub ahead of print: 13 Nov 2020].
12 Hashimoto DA, Rosman G, Rus D, *et al*. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70–6.

Supplemental Methods

Neural Network Architecture

The Neural Network architecture at the core of our algorithm is a DeepLabv3+ with a KSAC pooling layer [1] and a 101-layer ResNeSt backbone [2]. The network is trained for 72000 iterations with a batch size of 8 to minimize the cross-entropy loss with label smoothing. We set the initial learning rate for Stochastic Gradient-Descent to 1e-2 and polynomial decay over the training iterations. During training we randomly crop the input to equal height and width, apply horizontal and vertical flipping as well as slight alterations to brightness, hue, saturation, and contrast and add gaussian noise with a probability of 0.25.

Computer Vision Annotation Tool (CVAT)

In CVAT, each annotated instance represents a separate layer. These individual layers must be ordered from foreground to background, such that submucosal regions do not cover vessel annotations. Apparent conflicts in the annotations (Figure 1) are only present between the submucosal layer and the other annotated classes and are an artifact of the annotation process. It is more time-efficient to broadly annotate the submucosal region, omitting the subtle geometry of the knife or a small vessel and correct these conflicts with a post-processing by applying predefined ordering that always places the submucosal layer as background, in relation to the knife or vessel classes.

Image Annotation

Annotation of five categories within the training and test images was performed, including: 1) Submucosal vessels; 2) Submucosal layer; 3) Muscle layer, 4) Electrosurgical knife, 5) instrument shaft. Annotation was performed by expert endoscopists with an ESD experience of at least 500 procedures using the Computer Vision Annotation Tool (CVAT, doi: 10.5281/zenodo.4009388). The aim of annotation was to provide the ground truth for training and subsequent cross-validation or testing.

Training and Validation on still images

12 ESD- and 4 POEM-videos of about one hour duration per video were used for training and cross-validation. For the five-fold cross-validation, a total of 2012 frames were extracted from the videos. 453 further annotated frames from 9 ESD- and 2 POEM-videos were used for an additional performance test. These videos were not part of the training or cross-validation set. All images for training and validation were resized to a resolution of 512 x 640 pixels.

The individual folds are selected on a sequence level. Since all images from one sequence are only part of the validation set once, with varying amounts of images per sequence, each fold consists of a different number of training and validation data. Images were taken as screenshots from the ESD- and POEM-videos during the submucosal dissection stage and were selected to have a balanced distribution of the annotated classes. All procedures were performed at the University Hospital Augsburg using Olympus EVIS X1 CV-1500 series. Ethics approval for use of deidentified image and video material had been granted by the Ethics Committee of Ludwigs-Maximilians-Universität, Munich (Project Nr: 21-1216).

Validation on video

From three third-space endoscopic videos (1x rectal ESD; 1x esophageal ESD, 1x POEM) 31 video clips with special characteristics were extracted. Each video had to be 15 seconds to 100 seconds, within the first 5 seconds no vessel could be visible. To be included and regarded as relevant, a vessel had to have a diameter of at least 1mm [3] (reference: thickness of the electrosurgical knife shaft, Hook Knife J, Olympus, Tokyo, Japan). For two vessels to be counted separately within one clip, they had to have a space between each other of at least 3 mm (reference: thickness of the endoscopic instrument shaft). In Y-shaped vessels the same condition was applied for distance and length of the two arms. Vessels with a diameter of over 2mm were counted regardless of their distance to other vessels. From the 3 videos all vessels, which could be shown in clips according to these rules were extracted.

Hereby 27 videos containing a total of 52 predefined vessels were assembled. Four videos without vessels were also purposefully included in the test.

These videos were viewed frame by frame with AI overlay and for every positive measurement of a vessel it was determined visually if the measurement overlapped with a predefined vessel. For non-corresponding measurements it was determined visually, if a previously undetected vessel was visible, otherwise the measurement was counted as false positive. For analysis, false positive structures were counted.

Performance Measures

The algorithm's performance was evaluated by calculating the intersection over union (IoU) and Dice-Score. These metrics represent the percent overlap between expert annotation (ground truth) and the segmentation results of the algorithm. The IoU is the ratio between the correctly predicted area and the union of predicted and ground-truth regions. The Dice-Score is similar but puts a larger emphasis on the true positive region in the calculation. The pixel accuracy is computed for all classes at once and is the percentage of correct predictions among all predictions. All measures take values between 0 and 100 %. An IoU or Dice-Score of 0 % would mean no overlap between ground truth and AI prediction, while a Score of 100% would mean complete congruence between the two. If the prediction and ground truth have the same dimensions, but the prediction is shifted to the side such that only 50% of the prediction lies within the ground truth, the resulting IoU would be 33%. The degree of overlap that is satisfactory depends on the segmentation task in question, as in some circumstances, detection is more important than exact delineation.

$$IoU = TP / (TP + FP + FN)$$
$$Dice\ Score = 2\ TP / (2\ TP + FP + FN)$$
$$Pixel\ Accuracy = (TP + TN) / All$$

Abbreviations: TP = True Positives, FP = False Positives, TN = True Negatives, FN = False Negatives, All = all elements considered

To evaluate the model, we split the 16 video sequences into five cross-validation folds. The frames of a single video are either only present in the current training or the validation set. The presented validation metrics are calculated by accumulating the per-fold outputs in order to achieve one result for the whole validation set. The stated metrics are calculated from the fully trained model without early stopping on the best validation result.

In addition to the cross-validation results, we also demonstrate the performance on a separate test set that was strictly excluded during training. We applied the five

previously trained fold-specific models as an ensemble to the test data, such that the segmentation of a single testing image is the average output of the five fully trained models.
The VDR was determined as the number of correctly detected vessels divided by the number of predetermined vessels.

References:

1 Ye Huang, Qingping Wang, Wenjing Jia, Lu Yue, Xiangjian He. See More Than Once - Kernel-Sharing Atrous Convolution for Semantic Segmentation. arXiv preprint arXiv:190809443 2019.
2 Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, Alexander Smola. ResNeSt: Split-Attention Networks. arXiv preprint arXiv:200408955 2020.
3 Yoshida N, Naito Y, Kugai M, Inoue K, Wakabayashi N, Yagi N, *et al.* Efficient hemostatic method for endoscopic submucosal dissection of colorectal tumors. World J Gastroenterol 2010;**16**:4180-6.