

Automatic detection of tumor vessels in indeterminate biliary strictures in digital single-operator cholangioscopy



Authors

Pedro Pereira^{1,2,3}, Miguel Mascarenhas^{1,2,3}, Tiago Ribeiro^{1,2}, João Afonso^{1,2}, João P. S. Ferreira^{4,5}, Filipe Vilas-Boas^{1,2,3}, Marco P.L. Parente^{4,5}, Renato N. Jorge^{4,5}, Guilherme Macedo^{1,2,3}

Institutions

- 1 Department of Gastroenterology, São João University Hospital, Porto, Portugal
- 2 WGO Gastroenterology and Hepatology Training Center, Porto, Portugal
- 3 Faculty of Medicine of the University of Porto, Porto, Portugal
- 4 Department of Mechanical Engineering, Faculty of Engineering of the University of Porto, Porto, Portugal
- 5 INEGI – Institute of Science and Innovation in Mechanical and Industrial Engineering, Porto, Portugal

submitted 6.7.2021

accepted after revision 5.10.2021

Bibliography

Endosc Int Open 2022; 10: E262–E268

DOI 10.1055/a-1723-3369

ISSN 2364-3722

© 2022. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Corresponding author

Miguel José da Quinta e Costa de Mascarenhas Saraiva,
Rua Oliveira Martins 104, Porto, 4200-427, Portugal
Fax: +00351225509479
miguelmascarenhasaraiva@gmail.com

ABSTRACT

Background and study aims Indeterminate biliary strictures pose a significant clinical challenge. Dilated, irregular, and tortuous vessels, often described as tumor vessels, are frequently reported in biliary strictures with high malignancy potential during digital single-operator cholangioscopy (D-SOC). In recent years, the development of artificial intelligence (AI) algorithms for application to endoscopic practice has been intensely studied. We aimed to develop an AI algorithm for automatic detection of tumor vessels (TVs) in D-SOC images.

Patients and methods A convolutional neural network (CNN) was developed. A total of 6475 images from 85 patients who underwent D-SOC (Spyglass, Boston Scientific, Marlborough, Massachusetts, United States) were included. Each frame was evaluated for the presence of TVs. The performance of the CNN was measured by calculating the area under the curve (AUC), sensitivity, specificity, positive and negative predictive values.

Results The sensitivity, specificity, positive predictive value, and negative predictive value were 99.3%, 99.4%, 99.6% and 98.7%, respectively. The AUC was 1.00.

Conclusions Our CNN was able to detect TVs with high accuracy. Development of AI algorithms may enhance the detection of macroscopic characteristics associated with high probability of biliary malignancy, thus optimizing the diagnostic workup of patients with indeterminate biliary strictures.

Introduction

Diagnosis of biliary strictures (BS) is a clinical challenge, and although emerging technologies are being developed, establishing a correct diagnosis remains difficult in some patients. This is particularly relevant when the BS are located in the perihilar region or in the case of primary sclerosing cholangitis (PSC). Development of digital single-operator cholangioscopy (D-SOC) allowed direct visualization of the biliary epithelium in BS and

performing targeted biopsies. Data from clinical trials report an accuracy of 87% in visual diagnosis of BS [1]. In a recent meta-analysis, the diagnostic accuracy of D-SOC targeted biopsies was 85% [2]. More recent data from clinical trials report an overall accuracy of D-SOC of approximately 87% [1]. Furthermore, high success rates (96%) have also been reported in patients with PSC, in which multiple and fibrotic stenosis may limit cholangioscopy and a cholangioscopy-guided sample [3].

Despite the remarkable evolution of D-SOC, characterization of BS remains difficult. Indeed, diagnosing malignancy by visual impression has some limitations: accuracy is limited when evaluating extrinsic strictures (such as pancreatic cancer, gallbladder cancer or metastatic disease) compared to cholangiocarcinoma and irregular patterns of biliary mucosa may not represent malignancy [4]. In addition, pseudopolyp morphology and traumatic ulcers can be seen after stent removal, and even traumatic lesions due to the passage of the scope may be misinterpreted.

Multiple cholangioscopic findings suggestive of malignancy have been identified in the literature [5]. Indeed, visual classification of BS has shown to be sensitive in the prediction of malignant BS. Classifications for predicting the malignant potential of BS according to the presence of several morphologic features (intraductal masses or nodules, abnormal “tumor vessels” (TVs), papillary projections, ulceration and scarring) recently have been developed [5, 6]. Nevertheless, there is no consensual classification system for D-SOC morphologic findings and interobserver variability remains an issue. However, the most well-described cholangioscopic predictor of malignancy appears to be the presence of TV (tortuous and dilated vessels) [7]. These vessels represent the process of angiogenesis, a vital process in the progression of cancer, and can be identified by D-SOC in the superficial layers of the bile duct wall. Indeed, detection of irregular or spider vascularity on bile duct lesions during D-SOC evaluations accurately identifies biliary neoplastic lesions [8]. However, identification of TV in BS may be particularly difficult in the presence of chronic biliary tract inflammation, such as in PSC.

The introduction of artificial intelligence (AI) to routine endoscopic practice has been the focus of intense research over the last decade and has produced promising results [9, 10]. To date, the impact of AI algorithms, and particularly of convolutional neural networks (CNN), on the identification of macroscopic features of biliary lesions using D-SOC images has not been evaluated. The aim of this proof-of-concept study was to develop and validate a CNN-based model for automatic detection of TVs using D-SOC images.

Patients and methods

Population and study design

Subjects submitted to D-SOC between August 2017 and January 2021 at a single tertiary center (São João University Hospital, Porto, Portugal) were enrolled (n=85). Images obtained from these examinations were used for development, training, and validation of a CNN-based model for automatic identification of TVs and their distinction from benign biliary conditions.

This study was approved by the ethics committee of São João University Hospital (CE 41/2021) and respects the original and subsequent versions of the declaration of Helsinki. This study was retrospective and of non-interventional nature. Any information deemed to potentially identify the subjects was omitted. Each patient was assigned a random number in order to guarantee effective data anonymization. A team with Data Protection Officer certification confirmed the non-traceability of

data and conformity with the general data protection regulation.

Digital single-operator cholangioscopy procedure, definitions and data collection

All procedures were performed by two experienced endoscopists (P.P. and F.V.B.), using both the Spyglass DS and DSII (Boston Scientific Corp., Massachusetts, United States). Each of the researchers has performed more than 2000 ERCPs and 100 cholangioscopies. All procedures were performed with an Olympus TJF-160V or TFJ-Q180V duodenoscopes (Olympus Medical Systems, Tokyo, Japan). All obtained images were classified as showing a benign finding (comprising inflammatory vessels in BS of patients without evidence of biliary malignancy) or TV, if associated with histological evidence of malignancy. Identification of TV, defined as dilated/tortuous vessels and with spider vascularity resemblance, was performed independently by the two endoscopists (P.P. and F.V.B.). Final classification required consensus between both researchers. Images whose evaluation was not consensual were excluded from the datasets. A minimum of four biopsies were obtained during the procedure using the SpyBite or SpyBite Max biopsy forceps (Boston Scientific Corp., Marlborough, Massachusetts, United States), and the material fixed in formalin. The malignancy status of the BS was based on histopathology of biopsy or surgical specimens and no evidence/evidence of malignancy during a 6-month follow-up period.

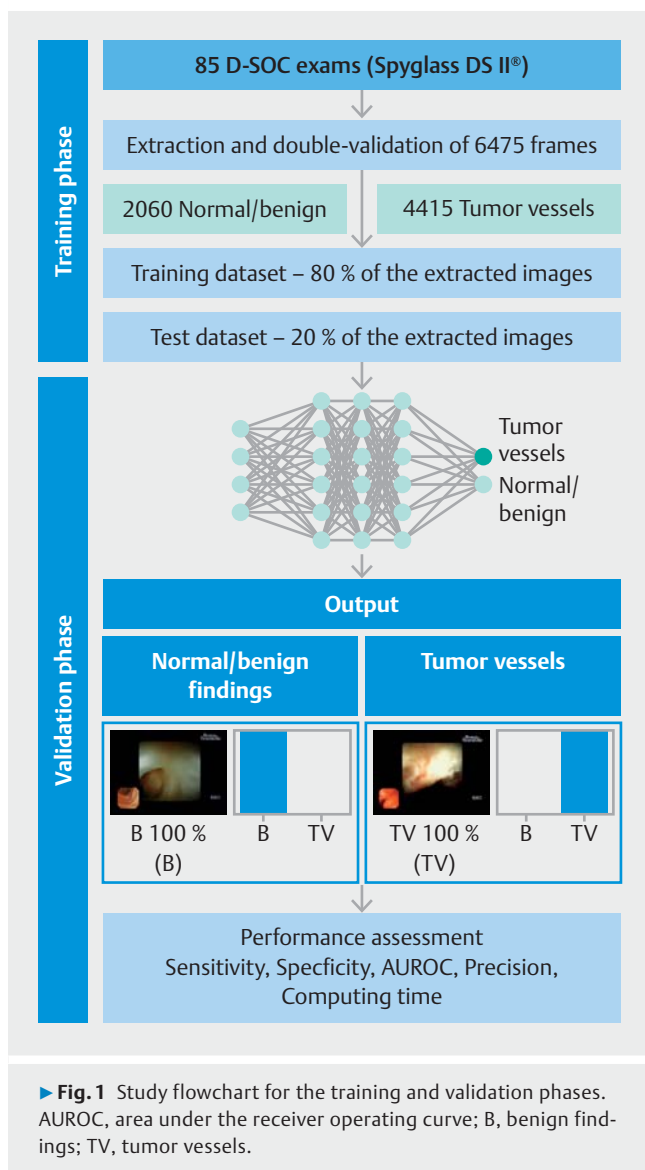
Development of the convolutional neural network

A deep learning CNN was developed for automatic identification of TV in D-SOC images. A total of 6475 images were collected (4415 TVs and 2060 showing benign findings). This pool of images was divided for constitution of training and validation datasets. The training dataset was composed of 80% of the extracted images (n=5180). The remaining 20% (n=1295) were used as the validation dataset for assessment of the performance of the CNN. The study flowchart is represented in ► **Fig. 1**.

The CNN was created using the Xception model with its weights trained on ImageNet. To transfer this learning to our data, we kept the convolutional layers of the model. We used Tensorflow 2.3 and Keras libraries to prepare the data and run the model. The analyses were performed with a computer equipped with a 2.1 GHz Intel Xeon Gold 6130 processor (Intel, Santa Clara, California, United States) and a double NVIDIA QuadroRTX 4000 graphic processing unit (NVIDIA Corp. California, United States).

Model performance and statistical analysis

A probability for each finding (either benign findings or TV associated with malignancy) was attributed by the CNN for every image (► **Fig. 2**). A higher probability demonstrated a greater confidence in the CNN prediction; the category with the highest probability was outputted as the CNN's classification. The classification provided by the CNN was compared to that of the endoscopist, which integrated data from visual impression (presence or absence of TV), histopathology and clinical



evolution. The classification provided by the endoscopists was considered the gold standard.

The baseline characteristics of the included patients are expressed as frequency and percentages for categorical variables, and median and interquartile range (IQR) for continuous variables. Categorical variables were compared using chi-square test whereas comparisons between continuous variables were made by the Mann-Whitney U test.

The primary outcome measures included sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively), accuracy, and area under the receiver operating characteristic curve (AUROC). In addition, the image processing performance of the network was determined by calculating the time required for the CNN to provide output for all images in the validation image dataset. Statistical analysis was performed using Sci-Kit learn v0.22.2 [11].

Results

Clinical and demographic data

Eighty-five patients underwent D-SOC between August 2017 and January 2021 and were included in the analysis. ▶ **Table 1** summarizes the baseline characteristics of patients. Forty-five patients (53%) were ultimately diagnosed with malignant stricture whereas the remaining 40 (47%) had benign disease. The median age was 65 (interquartile range 59 to 72 years) and 56 of 95 were male. A significant difference in the location of malignant or benign BS was found ($P < 0.01$). Malignant BS were most frequently located in the hepatic hilum (82.2%), whereas benign BS were most frequently intrahepatic. Malignant strictures were significantly longer than benign BS ($P < 0.01$). TVs were present in 43 of 85 of the included patients (50.6%): 41 of 45 patients with malignant BS (91.1%) and two of 40 patients had benign lesions (5.0%).

Construction of the network

Overall, 6475 frames were extracted for construction of the CNN: 4415 showed TVs and 2060 showed benign findings. The validation dataset (20%) comprised 1295 images, 829 having TVs and 466 showing benign findings. The accuracy of the CNN increased as data were repeatedly inputted into its multi-layer architecture (▶ **Fig. 3**).

Performance of the network

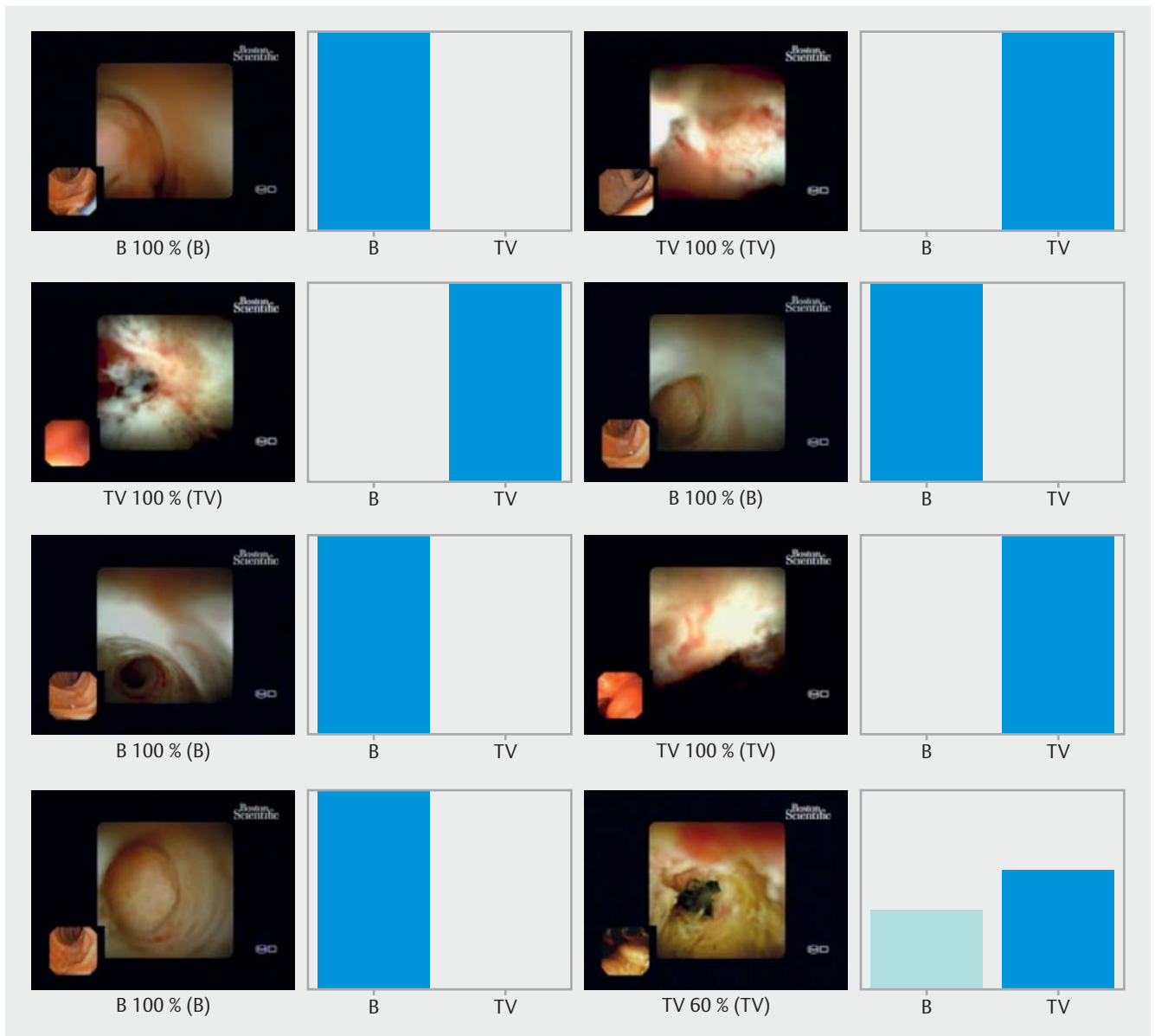
The performance of the CNN was evaluated using the trained CNN on the validation dataset. The confusion matrix between the trained CNN and final diagnosis is shown in ▶ **Table 2**. Overall, the model had a sensitivity and specificity of 99.3% and 99.4%, respectively, for detection of TVs associated with malignancy. The PPV and NPV were 99.6% and 98.7%, respectively. The overall accuracy of the network was 99.3%. The AUROC for detection of TVs was 1.00 (▶ **Fig. 4**).

Computational performance of the CNN

The CNN completed reading the validation dataset in 27 seconds. This translates into an approximate processing speed of 20 ms/image.

Discussion

Establishing a definitive diagnosis in patients with indeterminate BS is difficult due to the poor performance of routinely available diagnostic tools. Direct visualization of the lesion by D-SOC has improved the diagnostic yield in diagnosis of malignant biliary lesions [1]. Several macroscopic features have been linked to malignant BS [5]. TVs are one of the most common cholangioscopic findings in patients with known biliary neoplasia [12]. Nevertheless, detection of several macroscopic features associated with biliary malignancy has only achieved fair or moderate interobserver agreement. In fact, the interobserver agreement for detection of TVs was reported to be only fair ($\kappa = 0.26$) in a recent retrospective cohort study [5]. Poor specificity of macroscopic features and reproducibility between different observers, as well as the retrospective nature



► **Fig. 2** Output obtained during the training and development of the convolutional neural network. The bars represent the probability estimated by the network. The finding with the highest probability was outputted as the predicted classification. A blue bar represents a correct prediction. Red bars represent an incorrect prediction. B, benign biliary findings; TV, tumor vessels.

of studies, have limited development of a widely accepted D-SOC classification system for indeterminate BS [6, 13, 14].

In this pilot study, we report for the first time development of an AI model for detection of a single macroscopic feature for predicting the diagnosis of biliary malignancy. To our knowledge, this is the first study to evaluate the performance of a deep learning system for detection of TVs in patients with indeterminate BS. In addition, our CNN demonstrated high performance standards, with a sensitivity and specificity of 99%, an accuracy of 99%, and an AUROC of 1.00. Robles-Medrand et al. [8] recently evaluated use of neovascularity for identification of neoplastic bile duct lesions. Irregular TVs were present in 94% of patients with malignant lesions and 37% of patients with benign lesions. The vascularity pattern of lesions proved

useful to assess the malignancy status, achieving an accuracy of 80%, sensitivity of 94%, specificity of 63%, PPV of 75%, and NPV of 90%. The CNN developed by our group had higher performance levels compared with the results presented by Robles-Medrand and coworkers, showing a significantly enhanced specificity, PPV, and overall accuracy. The results of our proof-of-concept study build upon those presented by that group, demonstrating the potential gains in diagnostic performance from application of AI algorithms to D-SOC. Indeed, accurate automatic detection of macroscopic features associated with biliary malignancy, particularly TVs, may improve visual identification of areas with higher probability of malignancy, thus increasing the diagnostic rentability of cholangioscopy-targeted biopsies.

► **Table 1** Baseline characteristics of included patients.

	Overall (n = 85)	Malignant strictures (n = 45)	Benign strictures (n = 40)	P value
Sex				0.01
Male, n (%)	56 (65.9)	35 (77.8)	21 (52.5)	
Age				0.64
Years, median (IQR)	65 (59–72)	65 (58.5–71.5)	66 (60–74.5)	
Indication¹				<0.01
Biliary stricture, n (%)	47 (55.3)	32 (71.1)	15 (37.5)	
Filling defect, n (%)	9 (10.6)	–	9 (22.5)	
Indetermined CBD dilation, n (%)	19 (22.4)	3 (6.7)	16 (40.0)	
Extension of previously known CCa, n (%)	10 (11.8)	10 (22.2)	–	
Stricture location²				<0.01
CBD, n (%)	12 (16.9)	6 (13.3)	6 (23.1)	
Hilum, n (%)	46 (64.8)	37 (82.2)	9 (34.6)	
Intrahepatic, n (%)	13 (18.3)	2 (4.4)	11 (42.3)	
Stricture extension³,				<0.01
mm, median (IQR)	25.0 (15.0–37.0)	30.0 (20.0–38.0)	9.5 (4.8–6.3)	
Tumor vessels				<0.01
n (%)	43 (50.6)	41 (91.1)	2 (5.0)	
Adverse events⁴				0.70
Cholangitis, n (%)	7 (8.5)	4 (9.3)	3 (7.7)	
Pancreatitis, n (%)	14 (17.1)	9 (20.9)	5 (12.8)	
Perforation, n (%)	1 (1.2)	1 (2.3)		
Bacteremia, n (%)	1 (1.2)	1 (2.3)		

IQR, interquartile range; CCa, cholangiocarcinoma; CBD, common bile duct; CEA, carcinoembryonic antigen; CA 19–9, carbohydrate antigen 19–9; ERCP, endoscopic retrograde cholangiopancreatography.

¹ Based on previous imaging

² n = 26 for benign strictures

³ n = 27 for malignant strictures and n = 6 for benign strictures

⁴ n = 43 for malignant strictures and n = 39 for benign lesions

► **Table 2** Distribution of results of the validation dataset.

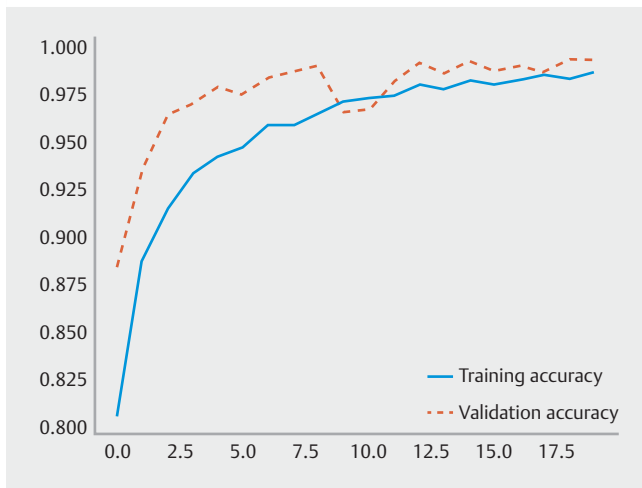
		Final diagnosis	
		Tumor vessels	Benign findings
CNN classification	Tumor vessels	823	3
	Benign findings	6	463

CNN, convolutional neural network.

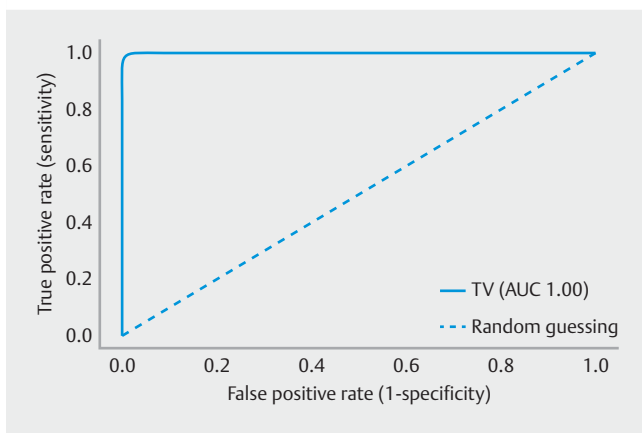
Tumor vessels were defined as dilated/tortuous vessels with spider vascularity that were associated with histological evidence of malignancy.

This study has limitations. First, it was retrospective and single-center. Second, our model analyzed still frames, and subsequent studies using full-length videos in real time are needed to accurately assess the clinical performance of these systems.

Nevertheless, considering the static and single nature of BS, our group is fairly confident of the future performance of our CNN in real-time D-SOC. Finally, this study was focused on evaluating a single endoscopic feature associated with malignancy.



► **Fig. 3** Evolution of accuracy of the convolutional neural network during training and validation phases, as the training and validation datasets were repeatedly inputted in the neural network.



► **Fig. 4** Receiver operating characteristic analysis of the network's performance in detection of malignant biliary strictures or benign biliary conditions. ROC, receiver operating characteristic; TV, tumor vessels.

Moreover, TVs may also occur in benign conditions, including IgG4 cholangiopathy and PSC.

Therefore, conjugating the automatic detection of multiple cholangioscopic features associated with malignancy would increase the significance of the results. Our work focused on detection of TVs, as they are one of the features most commonly associated with biliary malignancy. Future studies should focus on development of an algorithm incorporating several cholangioscopic patterns associated with biliary malignancy. Our group is currently working on models to address this limitation.

To the best of our knowledge, the impact of deep learning algorithms in identification of TVs in BS has not been evaluated. Our proof-of-concept model was highly accurate in detection of TVs. Further development of these systems may enable timely, accurate, and reproducible identification of TVs, thus optimizing the diagnostic process for patients with suspected biliary malignancy.

Acknowledgments

The authors acknowledge Fundação para a Ciência e Tecnologia (FCT) for supporting the computational costs related to this study through CPCA/A0/7363/2020 grant. This entity had no role in study design, data collection, data analysis, preparation of the manuscript, or the decision to publish.

Competing interests

The authors declare that they have no conflict of interest.

References

- [1] Gerges C, Beyna T, Tang RSY et al. Digital single-operator peroral cholangioscopy-guided biopsy sampling versus ERCP-guided brushing for indeterminate biliary strictures: a prospective, randomized, multicenter trial (with video). *Gastrointest Endosc* 2020; 91: 1105–1113
- [2] Wen LJ, Chen JH, Xu HJ et al. Efficacy and safety of digital single-operator cholangioscopy in the diagnosis of indeterminate biliary strictures by targeted biopsies: a systematic review and meta-analysis. *Diagnostics* 2020; 10: 666
- [3] Arnelo U, von Seth E, Bergquist A. Prospective evaluation of the clinical utility of single-operator peroral cholangioscopy in patients with primary sclerosing cholangitis. *Endoscopy* 2015; 47: 696–702
- [4] Chen YK, Pleskow DK. SpyGlass single-operator peroral cholangiopancreatography system for the diagnosis and therapy of bile-duct disorders: a clinical feasibility study (with video). *Gastrointest Endosc* 2007; 65: 832–841
- [5] Sethi A, Tyberg A, Slivka A et al. Digital Single-operator cholangioscopy (dsoc) improves interobserver agreement (IOA) and Accuracy for evaluation of indeterminate biliary strictures: The Monaco Classification. *J Clin Gastroenterol* 2020; doi:10.1097/mcg.0000000000001321
- [6] Robles-Medranda C, Valero M, Soria-Alcivar M et al. Reliability and accuracy of a novel classification system using peroral cholangioscopy for the diagnosis of bile duct lesions. *Endoscopy* 2018; 50: 1059–1070
- [7] Kim HJ, Kim MH, Lee SK et al. Tumor vessel: a valuable cholangioscopic clue of malignant biliary stricture. *Gastrointest Endosc* 2000; 52: 635–638
- [8] Robles-Medranda C, Oleas R, Sánchez-Carriel M et al. Vascularity can distinguish neoplastic from non-neoplastic bile duct lesions during digital single-operator cholangioscopy. *Gastrointestinal Endoscopy* 2021; 93: 935–941
- [9] Ding Z, Shi H, Zhang H et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; 157: 1044–1054. e1045
- [10] Hassan C, Spadaccini M, Iannone A et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; 93: 77–85. e76
- [11] Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: Machine learning in python. *J Mach Learn Res* 2011; 12: 2825–2830
- [12] Shah RJ, Rajman I, Brauer B et al. Performance of a fully disposable, digital, single-operator cholangiopancreatography. *Endoscopy* 2017; 49: 651–658

- [13] Sethi A, Doukides T, Sejjal DV et al. Interobserver agreement for single operator choledochoscopy imaging: can we do better? *Diagn Ther Endosc* 2014; 2014: 730731
- [14] Fukasawa Y, Takano S, Fukasawa M et al. Form-vessel classification of cholangioscopy findings to diagnose biliary tract carcinoma's superficial spread. *Int J Mol Sci* 2020; 21: 3311