

Automatic Identification of Papillary Projections in Indeterminate Biliary Strictures Using Digital Single-Operator Cholangioscopy

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

INTRODUCTION: Characterization of biliary strictures is challenging. Papillary projections (PP) are often reported in biliary strictures with high malignancy potential during digital single-operator cholangioscopy. 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 PP in digital single-operator cholangioscopy images.

METHODS: A convolutional neural network (CNN) was developed. Each frame was evaluated for the presence of PP. The CNN's performance was measured by the area under the curve, sensitivity, specificity, and positive and negative predictive values.

RESULTS: A total of 3,920 images from 85 patients were included. Our model had a sensitivity and specificity 99.7% and 97.1%, respectively. The area under the curve was 1.00.

DISCUSSION: Our CNN was able to detect PP with high accuracy. Future development of AI tools may optimize the macroscopic characterization of biliary strictures.

RESUMO

INTRODUÇÃO: A caracterização das estenoses biliares é uma tarefa desafiante. As projeções papilares (PP) são uma característica morfológica frequentemente relatada em durante a avaliação de estenoses biliares com alto potencial de malignidade por colangioscopia digital de operador único (CDOU). Nos últimos anos tem-se assistido a um desenvolvimento intenso de algoritmos de inteligência artificial (IA) para aplicação à prática endoscópica. O nosso objetivo foi desenvolver um algoritmo de IA para deteção automática de PP em imagens de CDOU.

MÉTODOS: Uma rede neural convolucional (RNC) foi desenvolvida. Cada imagem foi avaliada quanto à presença de PP. O desempenho da RNC foi medido pela área sob a curva (ASC), sensibilidade, especificidade, valores preditivos positivo e negativo.

RESULTADOS: Um total de 3920 imagens de 85 pacientes foram incluídas. O nosso modelo teve uma sensibilidade e especificidade de 99.7% e 97.1%, respetivamente. A ASC foi de 1.00.

DISCUSSÃO: A nossa RNC foi capaz de detectar PP com alta precisão. O desenvolvimento futuro de ferramentas de IA pode otimizar a caracterização macroscópica das estenoses biliares.

Ribeiro et al. *Clin Trans Gastroenterol*. 2021

Clinical and Translational
GASTROENTEROLOGY

Clinical and Translational Gastroenterology 2021;12:e00418. <https://doi.org/10.14309/ctg.0000000000000418>

¹Department of Gastroenterology, São João University Hospital, Alameda Professor Hernâni Monteiro, Porto, Portugal; ²WGO Gastroenterology and Hepatology Training Center, Porto, Portugal; ³Faculty of Medicine of the University of Porto, Alameda Professor Hernâni Monteiro, Porto, Portugal; ⁴Department of Mechanical Engineering, Faculty of Engineering of the University of Porto, Rua Dr. Roberto Frias, Porto, Portugal; ⁵INEGI—Institute of Science and Innovation in Mechanical and Industrial Engineering, Rua Dr. Roberto Frias, Porto, Portugal. **Correspondence:** Tiago Ribeiro, MD, MSc. E-mail: tiagofcriste@outlook.com.

*Tiago Ribeiro, MD, MSc, and Miguel Mascarenhas Saraiva, MD, MSc, contributed equally to this article.

Received July 12, 2021; accepted July 28, 2021; published online October 27, 2021

© 2021 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology

INTRODUCTION

The characterization of biliary strictures is a significant clinical challenge. In these patients, the main goal is to differentiate malignant from benign etiologies (1). The diagnosis is dependent on tissue sampling frequently obtained by endoscopic retrograde cholangiopancreatography–guided brush cytology or intraductal biopsies. However, the sensitivity of these methods is low, resulting in frequent falsely negative investigations for malignancy (2).

Digital single-operator cholangioscopy (D-SOC) enables high-resolution visualization of the bile duct lumen. Recent evidence has shown that D-SOC–guided biopsies increased the performance in differentiating malignant from benign lesions (3). Moreover, D-SOC allows for visual inspection of the bile duct, which has shown higher sensitivity and accuracy for the diagnosis of biliary malignancy compared with endoscopic retrograde cholangiopancreatography–guided tissue sampling (4). The presence of masses, nodules, dilated and tortuous vessels (tumor vessels), and papillary projections (PP) is associated with higher probability of malignancy (5). Nevertheless, the specificity of these findings is suboptimal, significant interobserver variability exists, and no validated scoring system exists to classify a biliary stricture as malignant based on visual examination (6).

Artificial intelligence (AI) tools for enhancement of endoscopic imaging have been the focus of intense research. However, the application of these technologies for automatic identification of morphologic features associated with biliary malignancy has

not been explored. We aimed to develop and validate a deep learning algorithm for automatic identification of PP in D-SOC images.

METHODS

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). All procedures were performed using Spyglass DS (Boston Scientific Corp., Marlboro, MA) by 2 experienced endoscopists (F.V.B. and P.P.). All obtained images were classified as showing a benign finding (including PP in patients without evidence of biliary malignancy) or PP, if these were associated with histological evidence of malignancy. The identification of PP required consensus between both researchers. A diagnosis of a benign biliary stricture was made in the case of negative histopathology of biopsy or surgical specimens and no evidence of malignancy during a 6-month follow-up period. This study was approved by the ethics committee of São João University Hospital (CE 41/2021).

A convolutional neural network (CNN) was developed for automatic identification of PP in D-SOC images. A total of 3,920 images were collected (1,650 PP and 2,270 showing benign findings). This pool of images was divided for constitution of training (80%) and validation (20%) data sets. The CNN was created using the Xception model with its weights trained on ImageNet. 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,

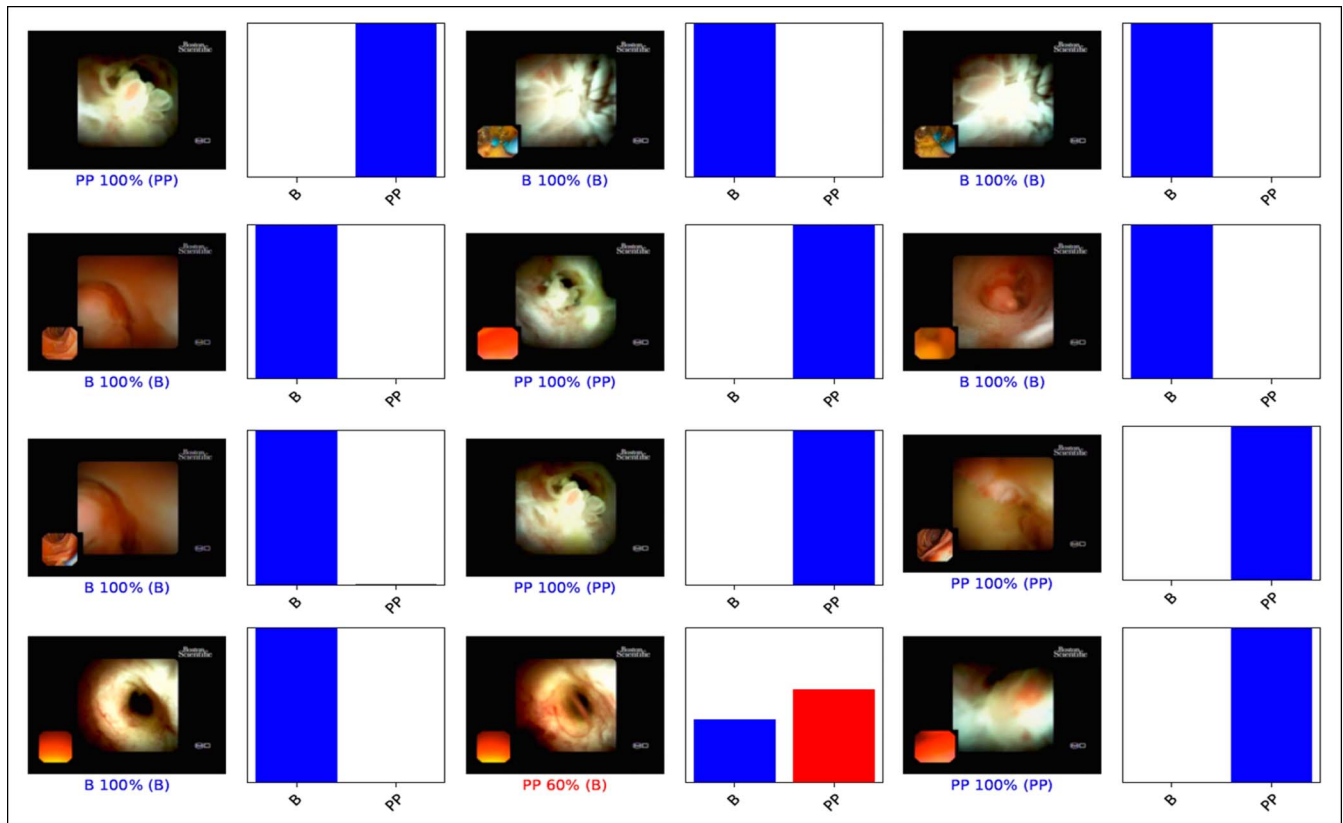


Figure 1. Output obtained during the training and development of the convolutional neural network. The bars represent the probability estimated by the network. The finding with the highest probability was outputted as the predicted classification. A blue bar represents a correct prediction. Red bars represent an incorrect prediction. B, benign biliary findings; PP, papillary projections.

Table 1. Distribution of results of the validation data set

	Expert's classification	
	Papillary projections	Benign findings
CNN's classification		
Papillary projections	329	13
Benign findings	1	441

CNN, convolutional neural network.

CA) and a double NVIDIA Quadro RTX 4000 graphic processing unit (NVIDIA Corp., Santa Clara, CA).

For each image, the CNN calculated the probability for each of category (Figure 1). 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 primary outcome measures included sensitivity, specificity, positive and negative predictive values, accuracy, and area under the receiver operating characteristic curve (AUC). Statistical analysis was performed using Sci-Kit learn v0.22.2 (7).

RESULTS

Construction of the network

A total of 3,920 frames were included. The validation data set (20%) comprised 784 images, 330 having PP, and 454 showing benign findings. The network showed increasing accuracy because data were being repeatedly inputted into the multi-layer CNN.

Performance of the network

The distribution of results is displayed in Table 1. Overall, the model had a sensitivity and specificity of 99.7% and 97.1% for the detection of PP. The positive predictive value and negative predictive value were 96.2% and 99.8%, respectively. The overall

accuracy of the network was 98.2%. The area under the receiver operating characteristic for detection of PP was 1.00 (Figure 2).

Computational performance of the CNN

The CNN completed the reading of the validation data set in 12 seconds. This translates into an approximate processing speed of 15 ms/image.

DISCUSSION

Digital cholangioscopy systems have a pivotal role in the evaluation of patients with suspected biliary malignancy. Direct visualization of the lesion allows for evaluation of its macroscopic characteristics, which has been shown to be highly sensitive for the diagnosis of malignant lesions (4). Several features have been associated with malignant bile strictures, including masses, tumor vessels, ulcerated lesions, and PP (5). Nevertheless, no macroscopic classification system for macroscopic classification of biliary strictures has been widely accepted, and significant interobserver variability exists in the description of these lesions (5,8). PP have been shown to correlate with the presence of malignancy (5,8,9). Indeed, Sethi et al. have found that PP were associated with malignancy in multivariate analysis (odds ratio 7.2, $P = 0.02$) (5). However, the interobserver agreement for the identification of this feature was suboptimal ($k = 0.43$) (5).

To date, the impact of deep learning algorithms in the identification of macroscopic features of biliary strictures has not been evaluated. The introduction of AI systems may allow the identification of these features, thus helping to predict the etiology of an indeterminate biliary stricture. Our proof-of-concept model was highly accurate in the detection of PP in malignant biliary strictures. Further development of these systems may allow a more precise evaluation of the macroscopic characteristics of a biliary lesion, thus reducing interobserver variability associated with human assessment.

This study has several limitations. First, it is a retrospective single-center study. Second, the number of frames included in

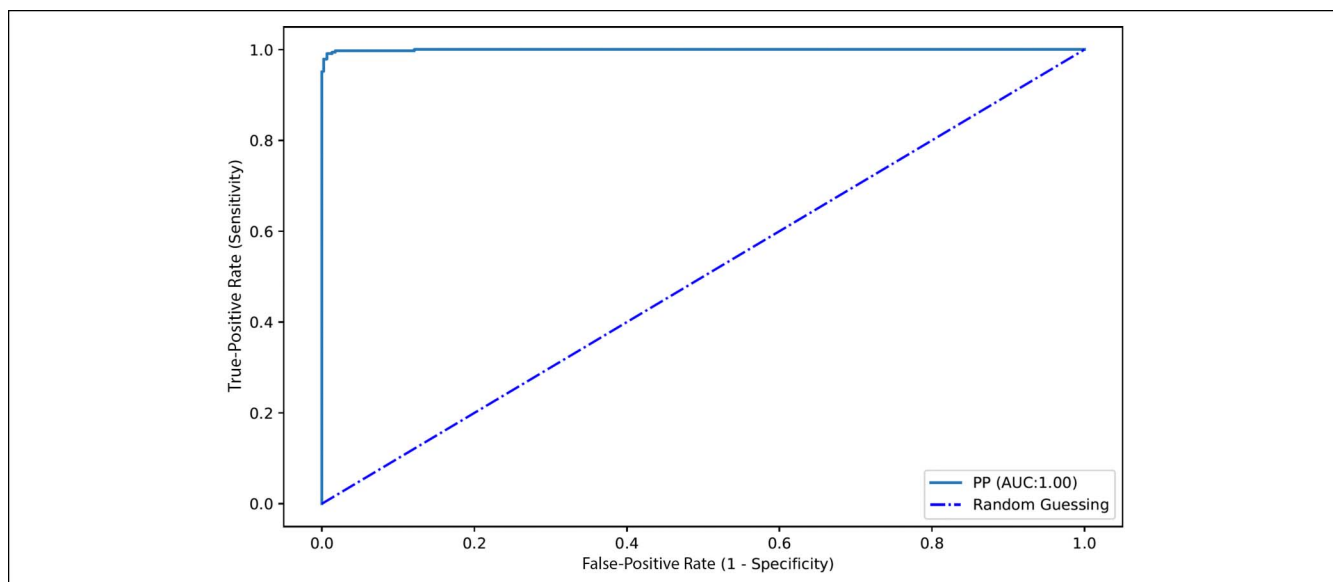


Figure 2. ROC analysis of the network's performance in the detection of malignant biliary strictures or benign biliary conditions. ROC, receiver operating characteristic; PP, papillary projections.

this study was small, hampering the generalizability of our results. Finally, our model analyzed still frames. Subsequent well-powered studies using full-length videos in real time are needed to accurately assess the clinical value of these systems.

In conclusion, the potential of AI algorithms for the investigation of patients with suspected biliary malignancy is vast. To the best of our knowledge, this is the first study to evaluate the potential of these systems for characterization of biliary lesions. Our proof-of-concept model lays the foundations for the development of deep learning algorithms for this subset of patients with the aim to optimize the diagnostic approach to these patients.

CONFLICTS OF INTEREST

Guarantor of the article: Tiago Ribeiro, MD, MSc.

Specific author contributions: T.R. and M.M.S.—study design, revision of D-SOC videos, image extraction and labeling and construction and development of the CNN, and data interpretation and drafting of the manuscript. J.A.—study design, revision of D-SOC videos, construction, and development of the CNN. J.P.S.F.—study design, construction and development of the CNN, and statistical analysis. P.P. and F.V.B.S.—equal contribution in study design, construction and development of the CNN, and data interpretation and drafting of the manuscript. M.P.L.P., R.N.J., and G.M.—study design and revision of the scientific content of the manuscript. All authors approved the final version of this manuscript.

Financial support: 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, and publishing decision.

Potential competing interests: None to report.

REFERENCES

1. Larghi A, Tringali A, Lecca PG, et al. Management of hilar biliary strictures. *Am J Gastroenterol* 2008;103:458–73.
2. Navaneethan U, Njei B, Lourdasamy V, et al. Comparative effectiveness of biliary brush cytology and intraductal biopsy for detection of malignant biliary strictures: A systematic review and meta-analysis. *Gastrointest Endosc* 2015;81:168–76.
3. Arvanitakis M. Digital single-operator cholangioscopy-guided biopsy for indeterminate biliary strictures: Seeing is believing? *Gastrointest Endosc* 2020; 91:1114–6.
4. 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;91:1105–13.
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.
6. Sethi A, Doukides T, Sejal DV, et al. Interobserver agreement for single operator choledochoscopy imaging: Can we do better? *Diagn Ther Endosc* 2014;2014:730731.
7. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Machine Learn Res* 2011;12:2825–30.
8. 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.
9. 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;50:1059–70.

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.