scientific reports

Check for updates

OPEN

Artificial intelligence for automatic diagnosis and pleomorphic morphological characterization of malignant biliary strictures using digital cholangioscopy

Miguel Mascarenhas[1,2,3,8]✉, Maria João Almeida[1,2], Mariano González-Haba[4], Belén Agudo Castillo[4], Jessica Widmer[5], António Costa[4], Yousef Fazel[5], Tiago Ribeiro[1,2,3], Francisco Mendes[1,2,3], Miguel Martins[1,2,3], João Afonso[1,2,3], Pedro Cardoso[1,2,3], Joana Mota[1,2], Joana Fernandes[6], João Ferreira[6,7], Filipe Vilas Boas[1,2,3], Pedro Pereira[1,2,3] & Guilherme Macedo[1,2,3]

Diagnosing and characterizing biliary strictures (BS) remains challenging. Artificial intelligence (AI) applied to digital single-operator cholangioscopy (D-SOC) holds promise for improving diagnostic accuracy in indeterminate BS. This multicenter study aimed to validate a convolutional neural network (CNN) model using a large dataset of D-SOC images to automatically detect and characterize malignant BS. D-SOC exams from three centers—Centro Hospitalar Universitário de São João, Porto, Portugal (*n* = 123), Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain (*n* = 18), and New York University Langone Hospital, New York, USA (*n* = 23)—were included. Frames were categorized based on histopathology. The CNN's performance in detecting tumor vessels, papillary projections, nodules, and masses was assessed. The dataset was split into 90% training and 10% validation sets. Performance metrics included AUC, sensitivity, specificity, PPV, and NPV. Analysis of 96,020 images from 164 D-SOC exams (50,427 malignant strictures and 45,593 benign findings) showed the CNN achieved 92.9% accuracy, 91.7% sensitivity, 94.4% specificity, 95.1% PPV, 93.1% NPV, and an AUROC of 0.95. Accuracy rates for morphological features were 90.8% (papillary projections), 93.6% (nodules), 93.2% (masses), and 78.1% (tumor vessels). AI-driven CNN models hold promise for enhancing diagnostic accuracy in suspected biliary malignancies. This multicenter study contributes diverse datasets to ongoing research, supporting further AI applications in this patient population.

Keywords Digital cholangioscopy, Artificial intelligence, Biliary strictures

Biliary strictures (BS) are frequently encountered in clinical practice, and accurately distinguishing between malignant and benign BS remains a diagnostic challenge. A biliary stricture with an etiology that cannot be ascertained after appropriate laboratory test, imaging study or endoscopic retrograde cholangiopancreatography (ERCP)-guided tissue biopsy is considered an indetermined BS[1]. Among these, a significant proportion represent malignant causes, portraying a poorer prognosis[2–6]. Malignant BS often stem from primary (cholangiocarcinoma) or secondary neoplasia with biliary tract extension (gallbladder, pancreatic, hepatocellular carcinoma, malignant lymph node compression)[7–9]. Conversely, approximately 30% of all BS are benign. Apart from iatrogenic causes,

[1]Department of Gastroenterology, Precision Medicine Unit, São João University Hospital, Porto, Portugal. [2]WGO Gastroenterology and Hepatology Training Center, Porto, Portugal. [3]Faculty of Medicine, University of Porto, Porto, Portugal. [4]Department of Gastroenterology, Hospital Universitario Puerta de Hierro Majadahonda, C/Joaquín Rodrigo, Majadahonda, Madrid 28220, Spain. [5]Department of Gastroenterology, New York University Langone Hospital, New York, USA. [6]Department of Mechanical Engineering, Faculty of Engineering, University of Porto, Porto, Portugal. [7]DigestAID—Digestive Artificial Intelligence Development, Rua Alfredo Allen n.o 455/461, Porto 4200-135, Portugal. [8]Gastroenterology Department Hospital de São João, Porto 4200-427, Portugal. ✉email: miguelmascarenhassaraiva@gmail.com

one must also consider biliary lithiasis, primary sclerosing cholangitis (PSC) and IgG4-related sclerosing cholangitis[10,11]. Adding to the complexity, malignancy may also develop in patients initially diagnosed with benign conditions such as PSC, underscoring the need for further evaluation if cholestasis worsens and/or a dominant stricture appears[12].

Following cross-sectional imaging, endoscopy retrograde cholangiopancreatography (ERCP) has historically served as the primary diagnostic modality in the management of patients with suspected BS. However, as in fluoroscopic gastrointestinal examinations, the cholangiographic indirect signs indicative of malignancy (surface irregularity, stricture length) are only moderately visible. As such, tissue sampling remains the preferred method to confirm a diagnosis of malignancy[12]. Nonetheless, ERCP-guided brush cytology or transpapillary intraductal biopsy have shown suboptimal diagnostic yields[13]. Indeed, a recent meta-analysis reported a sensitivity of 45% for brush cytology and 48% for ERCP-guided biopsies, which only slightly increased to 59% when both modalities were combined[14]. More recently, endoscopic ultrasound (EUS) has been applied for further investigation of biliary strictures, either by intraductal ultrasound (IDUS) or by an echoendoscope placed in the proximal duodenum. Despite its more favorable diagnostic performance comparatively to ERCP, the location of BS, particularly proximal BS, and the concern for needle track tumor seeding seems to affect its wider adoption[12,15].

Digital single-operator cholangioscopy (D-SOC) offers direct, high-resolution inspection of the biliary tree, enabling both a more accurate morphological characterization of BSs as well as the possibility of targeted biopsies[16,17]. In fact, a recent multicentric randomized trial demonstrated higher sensitivity of D-SOC for the visual identification of malignant strictures, compared to standard ERCP cholangiographic impression (96% vs. 67%, $p = 0.02$)[18]. Nonetheless, the specificity of the visual impression remains suboptimal (89%)[19]. Despite improvements on image resolution, interobserver variability of cholangioscopic visual features still exists even among advanced endoscopists. As a result, several classifications for visual prediction of BS malignancy have been tested, namely Monaco and Carlos Robles-Medranda classifications. The latest effort was the Mendoza criteria, which presented a high intraclass correlation for both neoplastic (0.90) and nonneoplastic (0.90) diagnosis and an overall diagnostic accuracy of 77% (range 64–88%)[20]. Despite that, a consensus for the visual diagnosis is not yet established and interobserver agreement remains poor[21].

Several morphological features present a particularly strong correlation with an increased malignancy risk, notably papillary projections and abnormal dilated tumor vessels[22,23]. However, once again, a lack of interobserver agreement regarding the former, and the challenging morphological identification of the latter in the presence of inflammatory disease, seems to add complexity to the visual assessment of cholangioscopies in cases of BS[24–26].

Consequently, targeted tissue sampling by cholangioscopy, as solely relying in visual impression of the area most likely affected by malignant involvement, demonstrates an unsatisfactory diagnostic accuracy, despite achieving significantly higher results when compared to ERCP-guided brushing (68% vs. 21%, $p < 0.01$)[18].

Convolutional neural networks (CNN) are human visual cortex inspired deep learning models tailored for image pattern recognition. Extensive literature has been published on the impact of these AI algorithms in various imaging-reliant procedures. Recently that impact has begun to be investigated on cholangioscopy images, with pilot studies, including some conducted by our group, already showing promising results[27,28].

In the light of the limitations in diagnostic approach of biliary strictures, our group sought to develop and validate a CNN model for automatic detection and differentiation between benign and malignant BSs in D-SOC images. Additionally, we aimed to assess the model's ability to identify significant morphological features of malignant BSs.

Materials and methods

Patient population and study design

D-SOC exams performed at three medical centers were collected from August 2017 until January 2023: Centro Hospitalar Universitário de São João (CHUSJ), Porto, Portugal; Hospital Universitario Puerta de Hierro Majadahonda (HUPHM), Madrid, Spain; and New York University Langone Hospital (NYULH), New York, USA. A total of 164 D-SOC exams (CHUSJ, $n = 125$; HUPHM, $n = 18$; NYULH, $n = 21$) were enrolled, from which 96,020 still-frame images were used for the development, training and validation phases of the CNN. The still-frame images were obtained during the exam, mainly through decomposition of the procedure videos into frames, using a VLC media player (VideoLAN, Paris, France, with selection of the segment of the video with indeterminate biliary stenosis and frame selection rate. Most commonly, 1 out of every 4 frames of the D-SOC video was included in the decomposition phase, to reduce the risk of model overfitting due to the presence of similar images. Subsequently, the region of the video containing the indeterminate biliary stricture was selected and the respective frames were evaluated with the AI model.

The study was performed following approval from the ethics committees of Centro Hospitalar Universitário de São João/Faculdade de Medicina da Universidade do Porto (CE 41/2021), Hospital Universitario Puerta de Hierro Majadahonda (PI 153/22) and New York University Langone Hospital (IRB 03845978/2023). This was a retrospective study conducted in accordance with the Declaration of Helsinki. To ensure patient confidentiality and anonymize the data, any potentially identifiable information was omitted, and each patient was assigned a random number. The non-traceability of the data and respect to the general data protection regulation (GDPR) was ensured by a team with a Data Protection Officer (DPO).

Digital-single operator cholangioscopy procedure and definitions

D-SOC exams included in the study were performed with the SpyGlassTM DS II system (Boston Scientific Corp., Marlborough, MA, USA) and LAN-EP-2612 system (Leinzett Medical, Zhejiang, China). The procedures were performed by expert gastroenterologists (P.P., F.V.B., M.G.-H., B.A.G and J.W.), each with prior experience of more than 2000 ERCPs and 100 cholangioscopies. The exams were performed using the Olympus TJF-160 V

or TFJ-Q180V duodenoscopes (Olympus Medical Systems, Tokyo, Japan), with biopsy specimens obtained using the SpyBite TM forceps (Boston Scientific Corp., Marlborough, MA, USA) under direct visual guidance, ensuring a minimum of four biopsies in all the study exams.

A total of 96,020 D-SOC biliary images were classified as either benign or malignant. Benign biliary findings typically included normal bile ducts, stone disease and benign BSs. A confirmed diagnosis of benign BS implied a negative histopathology (biopsy or surgical) with no evidence of malignancy after a 6-month follow up period[29]. Stone disease was diagnosed based on direct observation in the absence of other findings. A malignant diagnosis was confirmed by histopathology showing malignancy, obtained through D-SOC biopsy or other tissue sampling methods (namely brush cytology, fluoroscopic or endoscopic ultrasound- guided biopsy or surgical specimen).

Development of the convolutional neural network

We developed a deep learning-based CNN to automatically detect and differentiate malignant biliary strictures from benign biliary conditions, encompassing benign strictures, stone disease and normal bile ducts. A total of 96,020 frames were included: 50,427 images displaying malignant strictures, whereas the remaining 40,593 showing benign biliary conditions. The total data was separated into two sets: one for training, comprising 90% of the frames ($n = 86,418$), and the other for testing, consisting of 10% of the remaining images ($n = 9602$). The testing dataset was used to assess the model's performance. A graphical flowchart of the study design is shown in Fig. 1. Additionally, in a subset of exams ($n = 62$), a CNN was developed to detect morphological features associated with bile duct malignancy, notably "tumor vessels", "papillary projections", "nodules" and "masses". Tumor vessels were defined as abnormal, dilated, tortuous vessels ($n = 21,881$), while papillary projections ($n = 21,243$) were represented as finger-like projections.

The Resnet model served as construct to our CNN. ImageNet, a large-scale collection of images tailored for object recognition software development, was used to leveraging the weights between units. We retained its convolutional layers to transfer its learning to our model. Subsequently, the final fully connected layers were replaced by new ones fitted to the number of classes used to classify our endoscopic frames. Each of the two blocs used had an initially fully connected layer, followed by dropout layers with a drop rate of 0.1. After that, we added a dense layer whose size determined the number of classification groups: malignant or benign strictures. The model was trained with a learning rate of 0.0001, a batch size of 32, and a number of epochs of 10, using PyTorch framework. Performance evaluation was conducted on a computer equipped with a 2.1 GHz Intel® Xeon® Gold 6130 processor (Intel, Santa Clara, CA, USA) and dual NVIDIA Quadro® RTXTM 4000 graphics processing units (NVIDIA Corporate, Santa Clara, CA, USA).

Model performance and statistical analysis

The CNN's performance evaluation was conducted using an independent validation dataset (comprising 10% of all data). For each frame, the algorithm computed the probability of it being a malignant stricture or a benign biliary condition (Fig. 2). The final classification was assigned to the category with the highest probability, and it was subsequently compared against the corresponding histopathological assessment, considered as the gold standard. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy in differentiating malign strictures from benign biliary conditions were the primary outcomes assessed. Furthermore, we conducted ROC curve analysis and calculated the area under the ROC curve (AUROC) to assess the models' discriminatory ability. Additionally, the precision-recall curve and the area under the precision-recall curve (AUPRC) were utilized to evaluate the model's performance, considering potential data imbalances. Finally, we evaluated the algorithm's computational performance by measuring the time required for the CNN to process and generate output for all frames in the validation dataset. Statistical analysis was conducted using scikit-learn v0.22.2.

Results

Performance of the convolutional neural network

In total, 164 D-SOC exams were performed from August 2017 to January 2023, in 3 different medical centers – Portugal (Centro Hospitalar Universitário São João, Porto, $n = 123$), Spain (Hospital Puerta de Hierro Majadahonda, Madrid, $n = 18$) and United States of America (New York University Langone Hospital, New York, $n = 23$). We included 96,020 frames for the development of this CNN, of which 50,427 correspond to malignant strictures. The remaining 45,593 represented benign biliary conditions (benign strictures, stone disease and normal bile ducts). The baseline characteristics of the study population are presented in Table 1.

The model was trained and developed using 90% of the total dataset ($n = 86,418$). The remaining 9,602 images were used for assessment of algorithm's performance. Table 2 shows the confusion matrix between the CNN's predictions in validation set versus the histopathological characterization, considered the gold standard. Regarding the ability to detect and differentiate malignant strictures from benign conditions, the CNN revealed a sensitivity of 91.7%, a specificity of 94.4%, and an accuracy of 92.9%. PPV and NPV were 94.8% and 91.1%, respectively. The model's metrics were associated with an image reading rate of 365 frames per second. The model's AUROC and AUPRC were both 0.95, as shown in Fig. 3

Detection of morphological characteristics associated with biliary malignancy

The CNN's performance for the detection of morphological features associated with malignancy of the biliary tract (tumor vessels, papillary projections, nodules and masses) were also assessed on a subset of exams of patients with malignant biliary strictures. In this context, the CNN included 21,881 images of tumoral vessels, 21,243 of papillary projections, 2081 of nodules and 2444 images of masses (Table 3).

Heatmaps were generated for the identification of features contributing to the predictions of the CNN (Fig. 4). Regarding tumor vessel detection, the CNN sensitivity and specificity were 87.8% and 62.3%, respectively, with

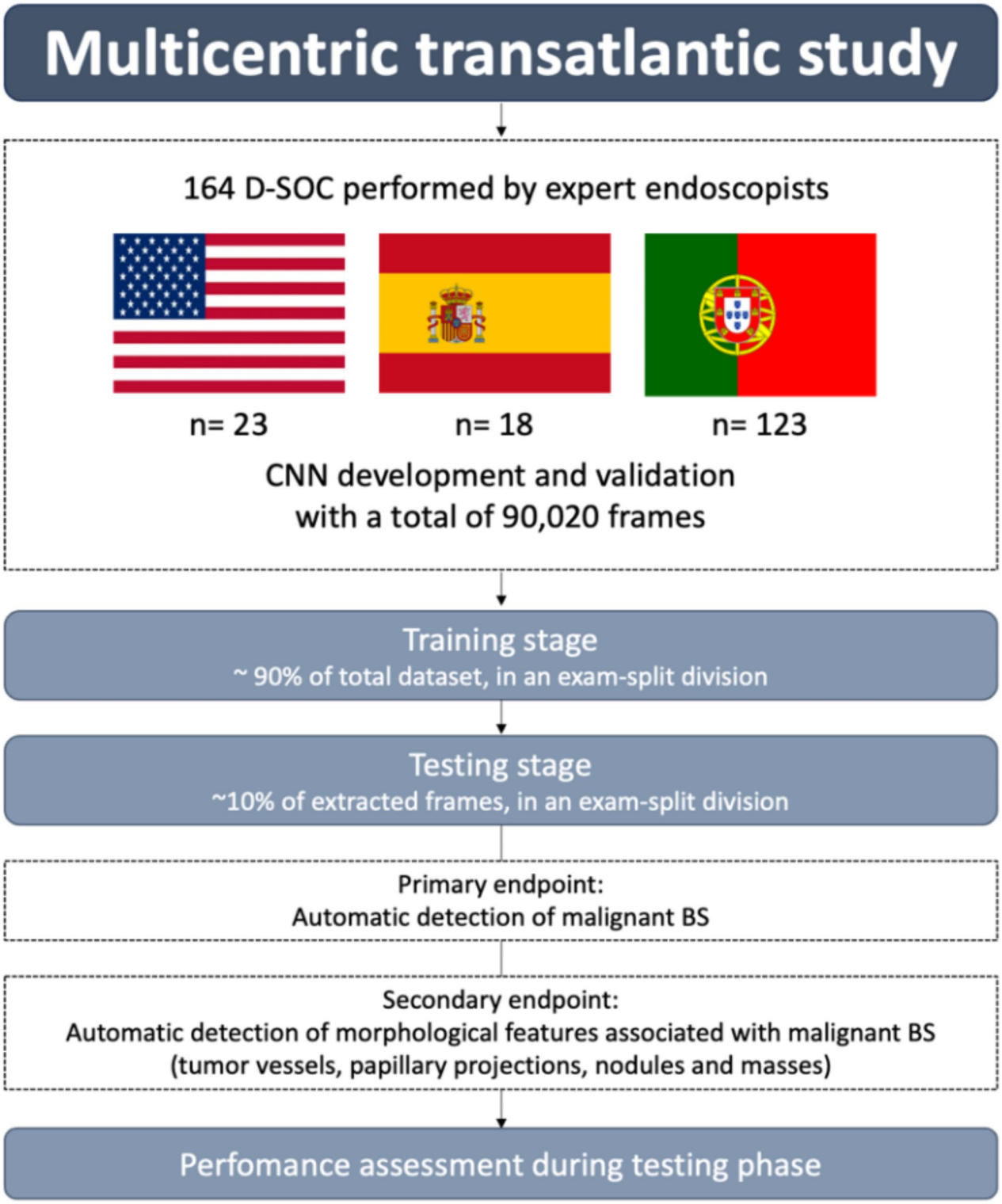

Fig. 1. Study flowchart for the training and testing stages.

an accuracy of 78.1%. In terms of papillary projection identification, the model's sensitivity, specificity and accuracy were 59.8%, 97.4% and 91.2%, respectively. In the case of nodules, the model's sensitivity, specificity and accuracy was 95.1%, 91.9% and 93.6%. Regarding masses, our model attained a sensitivity of 92.8%, a specificity of 93.5%, and an accuracy of 93.2%. The AUC for the detection of each morphologic feature is shown in Fig. 5.

Discussion.

Obtaining a timely and definitive diagnosis in indeterminate BS holds paramount importance in tailoring treatments for each patient[3,6,9]. Nonetheless, despite the advances in cross-sectional imaging and pancreatobiliary

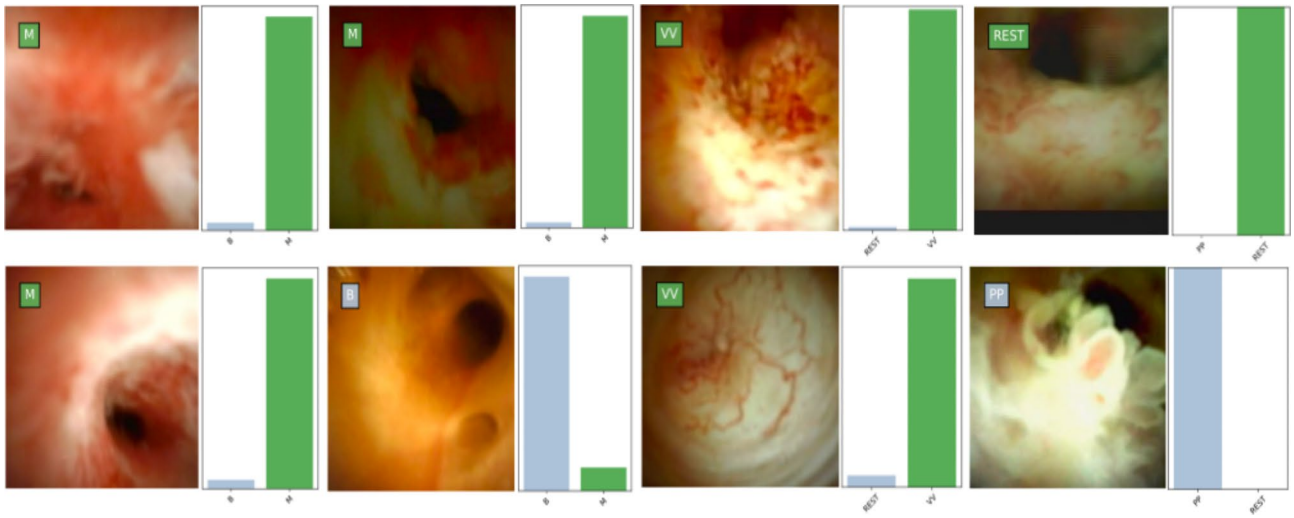

Figure 2. Output obtained during the training and development of the convolutional neural network. The bars represent the probability estimated by the network. The finding with the highest probability was output as the predicted classification. The bar represents a correct prediction. B – benign biliary findings; M – malignant stricture; PP – papillary projections; REST - other; VV – tumor vessels.

endoscopy, achieving a precise diagnosis remains challenging, mainly due to frequently inconclusive tissue sampling. The introduction of D-SOC has notably enhanced pancreatobiliary diagnostics, particularly evident in the accuracy of visually assessing significant biliary lesions, with several studies suggesting an estimated sensitivity of visual impression to be around 90%[30]. However, the diagnosis of malignancy solely based on visual impression is hindered by suboptimal specificity and accuracy[31,32]. Moreover, the lack of a universally accepted imaging classification system and poor interobserver agreement on existing systems further increases the challenges in this field. Nevertheless, beyond the possibility of macroscopically evaluate the biliary mucosa, cholangioscopy offers the advantage of performing targeted biopsies of ductal lesions under direct visualization. Recent systematic review has estimated the sensitivity of D-SOC-guided biopsy to be around 74%[33]. Although limited data exist on direct comparison between ERCP and D-SOC-guided tissue sampling, studies already demonstrated significantly higher sensitivity of D-SOC-based biopsies compared to those obtained during ERCP procedures[18].

However, when dealing with lesions with a notable malignant potential and a poorer prognosis if left untreated, a missing rate as high as 10% for D-SOC on visual inspection or targeted biopsies is far from satisfactory[19]. Considering these constrains, integrating AI technology into D-SOC holds promise for effectively managing these challenges. Significant research efforts have focused on the identification of malignant strictures on D-SOC images, with very promising results already achieved. By integrating visual features strongly linked to a higher probability of malignancy, these AI systems can assist in identifying regions where suspected malignant lesions are located, which, in turn, has the potential to further enhance the diagnostic yield of D-SOC-guided biopsies.

Our model had a binary endpoint: not only categorize biliary strictures into benign or malignant, but also to identify the morphological features associated with an increased risk of malignancy, namely tumor vessels, papillary projections, nodules and masses. Overall, the model demonstrated exceptional sensitivity, specificity and accuracy on both tasks, paving once more the way for a clinical adoption of these systems in the diagnosis and management of biliary strictures.

The study has some favorable points worth to mention. First, to the best of our knowledge, it is the first transatlantic multicentric deep learning algorithm to be developed for analysis of D-SOC images retrieved from three high-volume referral centers. Second, the robust dataset of more than 90 000 images of patients with BSs require the malignant diagnosis to be unequivocal biopsy-proven. Thirdly, the present study can be understood as a continuum of previously published research on these grounds, where an enrich dataset, not only in quantity but also in variety, further expands the evidence supporting the application of these algorithms as well as minimizes any demographic biases that may exist due to differences in the incidence of biliary disease in the investigated population. Fourthly, similar to other medical AI applications, the interoperability concern was partially addressed, as our model was developed and tested on more than one D-SOC platform, smoothing its generalization into clinical practice. Additionally, our model also generated results in the form of heatmaps, therefore contributing to the aim of explainability in the application of these AI algorithms. These heatmaps identify the area of the image most contributing to the output of the model. It is expected that, with mounting evidence, the generation of these solutions may enable the performance of targeted biopsies for regions identified by the algorithm as most probably harboring malignancy, therefore contributing to increase the diagnostic yield of D-SOC-guided biopsies. Nevertheless, to this date, studies on these explainable AI techniques for cholangioscopy have not been performed, and correlation with histopathological findings will need to be demonstrated. Finally, it is also worth to mention that the study is compliant with the FAIR principles for scientific data investigations: it is Findable as data is uniquely identified, Accessible for a considerable large

	Patients ($n = 164$)
Age, years (SD)	65.5 (10.0)
Sex	
Female, n (%)	56 (34.2%)
Male, n (%)	108 (65,8%)
Study Center	
Centro Hospitalar Universitário São João, Portugal, n (%)	123 (75.0%)
New York University Langone Hospital, USA, n (%)	23 (14.0%)
Hospital Universitario Puerta de Hierro Majadahonda, Spain, n (%)	18 (11,0%)
D-SOC Device	
SpyGlass™ DS II system (Boston Scientific®)	158 (96.3%)
LAN-EP-2612 system (Leinzett Medical®)	6 (3.7%)
Indication	
Biliary duct stricture, n (%)	68 (41.5%)
Jaundice, n (%)	54 (32.9%)
Biliary duct dilation, n (%)	26 (15.8%)
Abdominal pain, n (%)	20 (12.2%)
Pruritus, n (%)	15 (9.1%)
Biliary duct lithiasis, n (%)	13 (7.9%)
Diagnosis	
Benign, n (%)	67 (40.8)
Malignant, n (%)	97 (59.2)
Morphological features of malignant biliary strictures	
Tumor-like vessels, n (%)	61 (62.9)
Masses, n (%)	18 (18.6)
Papillary projections, n (%)	16 (16.5)
Nodules, n (%)	14 (14.4)
Biliary stricture topography	
Hilar, n (%)	61 (44.2)
Common bile duct, n (%)	47 (34.0)
Intrahepatic, n (%)	30 (21.8)
Previous biliary stent	
Yes, n (%)	31 (18.9)
No, n (%)	133 (81.1)

Table 1. Baseline characteristics. SD – standard deviation. USA – United States of America.

		Final diagnosis	
		Malignant	Benign
CNN classification	Malignant	4624	255
	Benign	418	4304

Table 2. Confusion matrix of the automatic detection versus final diagnosis, CNN—convolutional neural network; malignant—malignant biliary strictures; Benign—normal bile ducts or benign biliary findings.

population sample, Interoperable across different systems and Reusable as it represents a continuum of the work previously developed.

However, some limitations must be acknowledged. First, despite the relatively large dataset used, it was retrospectively conducted. Therefore, external clinical validation of this algorithm requires prospective studies with more robust statistical relevance. Secondly, benign strictures were defined according to existing position statements on the issue, which based on limited evidence. This adoption of this period follows previous studies which described that most neoplasms (79%) identified in patients with atypical cells in brush cytology were diagnosed on the first 6 months of follow-up[34]. Nevertheless, the authors acknowledge that a non-negligible fraction of patients are diagnosed after the 6-month period, particularly in the 7–12-month period[34]. Therefore, future studies on this subject should aim to include larger follow-up times to improve the accuracy in the characterization of benign biliary strictures. Third, despite the use of two cholangioscopy systems, the increasing number of commercial solutions requires subsequent expansions of the algorithm to accommodate the several platforms therefore ensuring interoperability. Thus, the results of this study will need to be confirmed on the

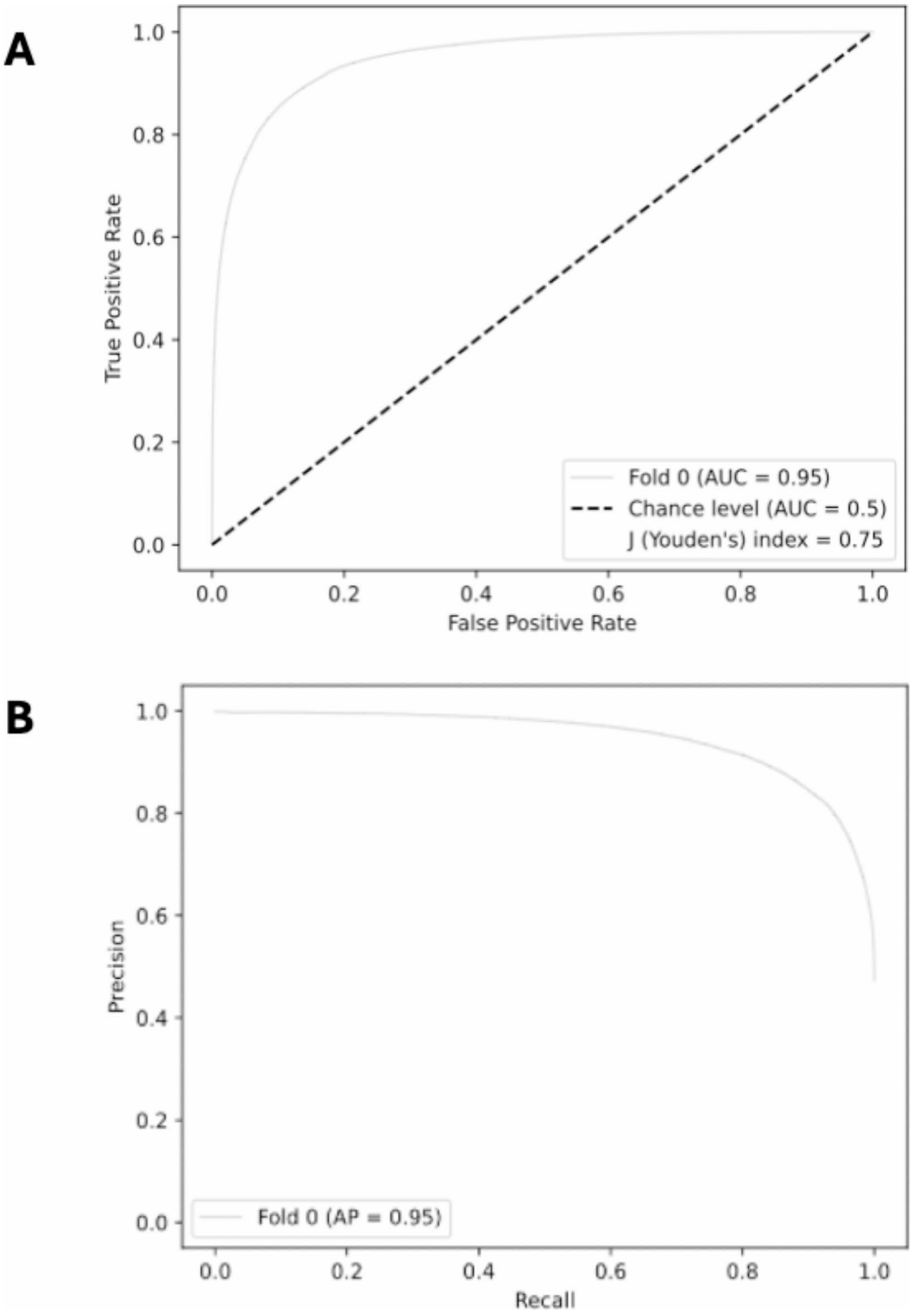

Fig. 3. Receiver operating characteristic (**A**) and precision-recall (**B**) analysis of the network's performance in the detection of malignant biliary strictures or benign biliary conditions.

various existing platforms before validation. the study only evaluated CNN at still frame level and, despite achieving good performance results, equivalent performances with real-time full-length video might not be ensured. Therefore, more conclusive data should be obtained through well powered prospective studies evaluating this solution, before it can be introduced to real-life clinical practice. Moreover, overfitting problems cannot be totally excluded, even though efforts were made to mitigate that risk. Finally, careful consideration of

	Tumor vessels	Papillary projections	Nodules	Masses
Training set	15,316	19,118	1873	2200
Testing set	6565	2125	208	244
Total	21,881	21,243	2081	2444

Table 3. Distribution of frames between training and testing sets with presence of tumor vessels papillary projections, nodules and masses (features more commonly associated with malignancy).

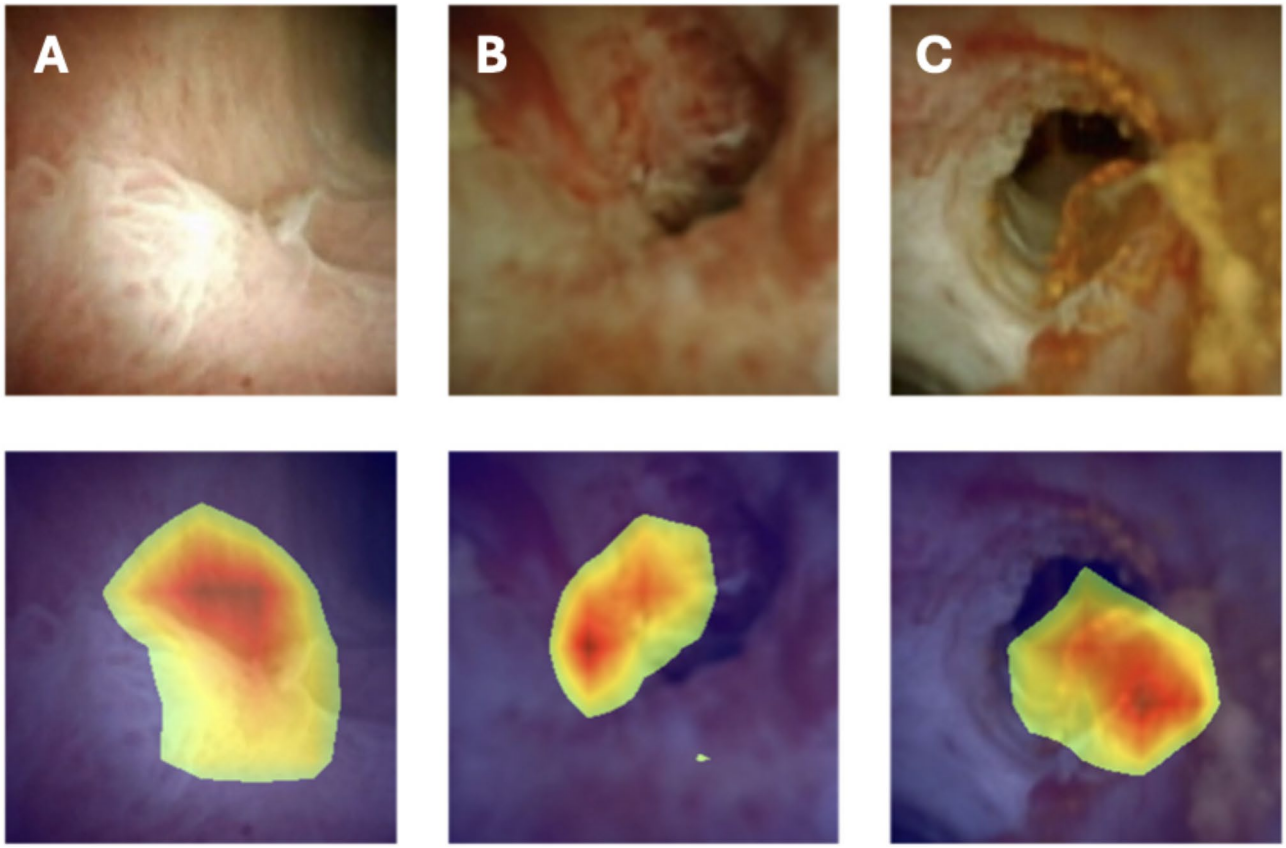

Fig. 4. Heatmaps generated, identifying characteristics typically associated with malignancy, particularly papillary projections (**A**) and anormal vessels (**B** and **C**).

legal and ethical implications needs to be taken to guarantee the required transparency while maximizing the overall performance.

Concluding, the exponential growth of AI is expected to continue with no exception to pancreatobiliary procedures. This multicentric transatlantic study was developed and validated for automatic differentiation between malignant and benign biliary disorders, using an enrich dataset composed by D-SOC frames from three experienced centers in this field. The optimistic performance metrics represent an ongoing progress on the work developed in recent years on these grounds, with the ultimate goal of enhancing the clinical outcomes for individuals suspected of having biliary malignancy.

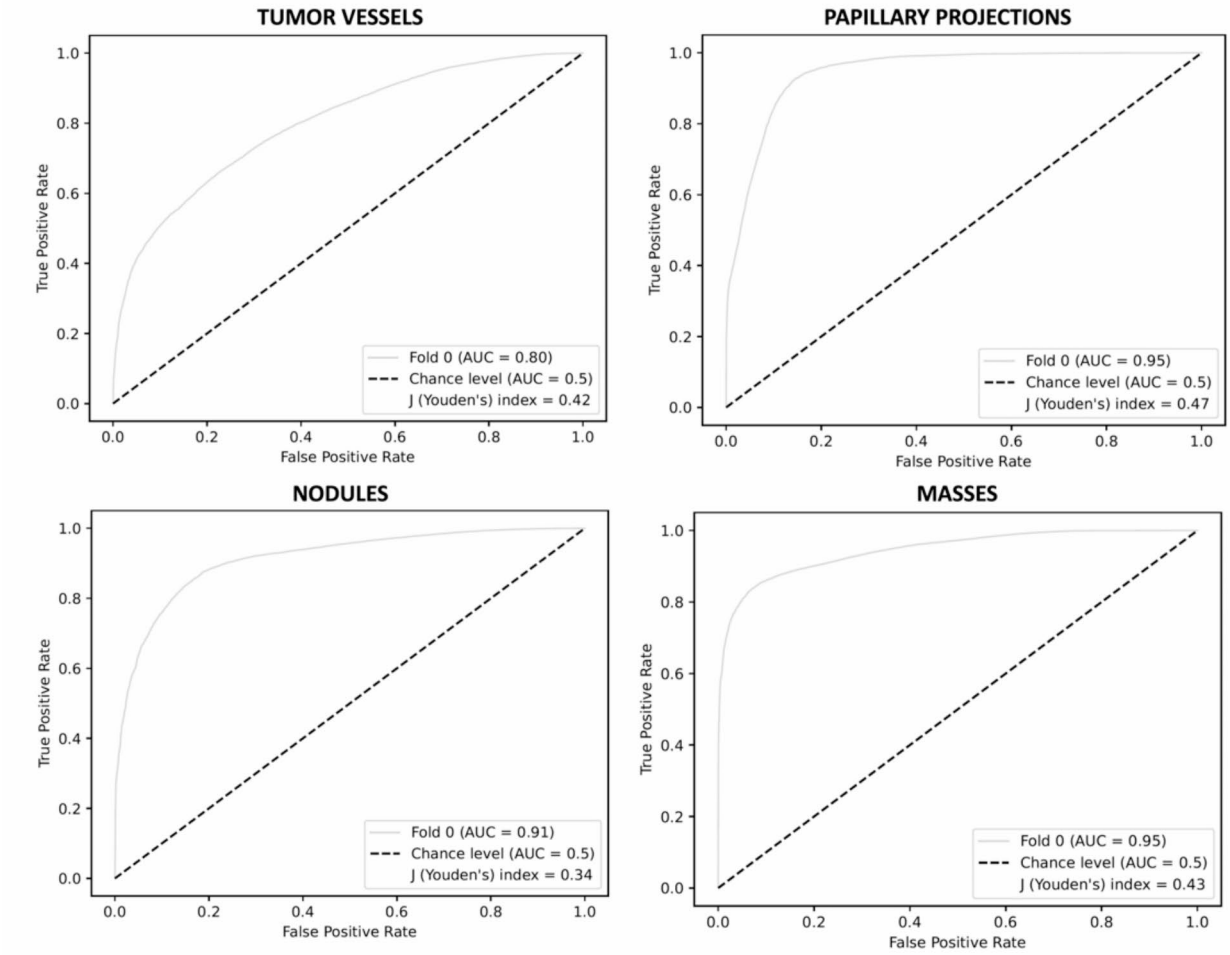

Fig. 5. Receiver Operating Characteristics (ROC) curve analysis of the network's performance in the detection of morphological characteristics of malignancy, namely tumoral vessels, papillary projections, nodules and masses.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 27 July 2024; Accepted: 17 January 2025
Published online: 14 February 2025

References

1. Ayoub, F. & Othman, M. O. Guidelines on cholangioscopy for indeterminate biliary strictures: one step closer to consensus. *Hepatobiliary Surg. Nutr.* **12**, 776 (2023).
2. Sahin, T. K., Rizzo, A., Aksoy, S. & Guven, D. C. Prognostic significance of the Royal Marsden Hospital (RMH) score in patients with Cancer: a systematic review and Meta-analysis. *Cancers (Basel)*. 16. https://doi.org/10.3390/cancers16101835 (2024).
3. Rizzo, A. & Brandi, G. Pitfalls, challenges, and updates in adjuvant systemic treatment for resected biliary tract cancer. *Expert Rev. Gastroenterol. Hepatol.* **15**, 547–554. https://doi.org/10.1080/17474124.2021.1890031 (2021).
4. Guven, D. C. et al. The association between albumin levels and survival in patients treated with immune checkpoint inhibitors: a systematic review and meta-analysis. *Front. Mol. Biosci.* **9**, 1039121. https://doi.org/10.3389/fmolb.2022.1039121 (2022).
5. Brandi, G. et al. Activated FGFR2 signalling as a biomarker for selection of intrahepatic cholangiocarcinoma patients candidate to FGFR targeted therapies. *Sci. Rep.* **14**, 3136. https://doi.org/10.1038/s41598-024-52991-8 (2024).
6. Rizzo, A. & Brandi, G. Neoadjuvant therapy for cholangiocarcinoma: a comprehensive literature review. *Cancer Treat. Res. Commun.* **27**, 100354. https://doi.org/10.1016/j.ctarc.2021.100354 (2021).
7. Larghi, A., Tringali, A., Lecca, P. G., Giordano, M. & Costamagna, G. Management of hilar biliary strictures. *Official J. Am. Coll. Gastroenterol.| ACG.* **103**, 458–473 (2008).
8. Pimpinelli, M., Makar, M. & Kahaleh, M. Endoscopic management of benign and malignant hilar stricture. *Dig. Endoscopy*. **35**, 443–452 (2023).
9. Rizzo, A. et al. How to choose between Percutaneous Transhepatic and endoscopic biliary drainage in malignant obstructive jaundice: an updated systematic review and Meta-analysis. *Vivo* **34**, 1701–1714. https://doi.org/10.21873/invivo.11964 (2020).

10. Tummala, P., Munigala, S., Eloubeidi, M. A. & Agarwal, B. Patients with obstructive jaundice and biliary stricture ± mass lesion on imaging: prevalence of malignancy and potential role of EUS-FNA. *J. Clin. Gastroenterol.* **47**, 532–537 (2013).
11. Paranandi, B. & Oppong, K. W. Biliary strictures: endoscopic assessment and management. *Frontline Gastroenterol.* **8**, 133–137 (2017).
12. Tang, R. S. Endoscopic evaluation of indeterminate biliary strictures: Cholangioscopy, endoscopic ultrasound, or both? *Digestive Endoscopy* (2023).
13. Burnett, A. S., Calvert, T. J. & Chokshi, R. J. Sensitivity of endoscopic retrograde cholangiopancreatography standard cytology: 10-y review of the literature. *J. Surg. Res.* **184**, 304–311 (2013).
14. Navaneethan, U. 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.* **81**, 168–176 (2015).
15. Cheon, Y. K. Intraductal ultrasonography for biliary strictures. *Clin. Endoscopy*. **56**, 164 (2023).
16. Pereira, P. et al. How SpyGlass™ may impact endoscopic retrograde cholangiopancreatography practice and patient management. *GE-Portuguese J. Gastroenterol.* **25**, 132–137 (2018).
17. Mauro, A. et al. The role of Cholangioscopy in biliary diseases. *Diagnostics* **13**, 2933 (2023).
18. Gerges, C. 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.* **91**, 1105–1113 (2020).
19. Jang, S., Stevens, T., Kou, L., Vargo, J. J. & Parsi, A. M. Efficacy of digital single-operator cholangioscopy and factors affecting its accuracy in the evaluation of indeterminate biliary stricture. *Gastrointest. Endosc.* **91**, 385–393 (2020). e381.
20. Kahaleh, M. et al. Digital single-operator cholangioscopy interobserver study using a new classification: the Mendoza classification (with video). *Gastrointest. Endosc.* **95**, 319–326 (2022).
21. Sethi, A. et al. Interobserver agreement for single operator choledochoscopy imaging: can we do better? *Diagn. Therapeutic Endoscopy* **2014** (2014).
22. Robles-Medranda, C. et al. Reliability and accuracy of a novel classification system using peroral cholangioscopy for the diagnosis of bile duct lesions. *Endoscopy* **50**, 1059–1070 (2018).
23. Fukasawa, Y. et al. Form-vessel classification of cholangioscopy findings to diagnose biliary tract carcinoma's superficial spread. *Int. J. Mol. Sci.* **21**, 3311 (2020).
24. Angsuwatcharakon, P. et al. Consensus guidelines on the role of cholangioscopy to diagnose indeterminate biliary stricture. *HPB* **24**, 17–29 (2022).
25. Kulpatcharapong, S., Pittayanon, R., Kerr, S. J. & Rerknimitr, R. Diagnostic performance of different cholangioscopes in patients with biliary strictures: a systematic review. *Endoscopy* **52**, 174–185 (2020).
26. El Bacha, H. et al. Identification of endoscopic predictors of biliary malignancy during digital cholangioscopy. *Dig. Endoscopy*. **34**, 1224–1233 (2022).
27. Saraiva, M. M. et al. Artificial intelligence for automatic diagnosis of biliary stricture malignancy status in single-operator cholangioscopy: a pilot study. *Gastrointest. Endosc.* **95**, 339–348 (2022).
28. Saraiva, M. M. et al. Deep Learning for Automatic diagnosis and morphologic characterization of malignant biliary strictures using Digital Cholangioscopy: a multicentric study. *Cancers* **15**, 4827 (2023).
29. Angsuwatcharakon, P. et al. Consensus guidelines on the role of cholangioscopy to diagnose indeterminate biliary stricture. *HPB (Oxford)*. **24**, 17–29. https://doi.org/10.1016/j.hpb.2021.05.005 (2022).
30. Navaneethan, U. et al. Digital, single-operator cholangiopancreatoscopy in the diagnosis and management of pancreatobiliary disorders: a multicenter clinical experience (with video). *Gastrointest. Endosc.* **84**, 649–655 (2016).
31. Sun, X. et al. Is single-operator peroral cholangioscopy a useful tool for the diagnosis of indeterminate biliary lesion? A systematic review and meta-analysis. *Gastrointest. Endosc.* **82**, 79–87 (2015).
32. Sethi, 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.* **56**, e94–e97 (2022).
33. Wen, L. J., Chen, J. H., Xu, H. J., Yu, Q. & Liu, K. 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* **10**, 666 (2020).
34. Navaneethan, U. et al. Predictors for detection of cancer in patients with indeterminate biliary stricture and atypical cells on endoscopic retrograde brush cytology. *J. Dig. Dis.* **15**, 268–275. https://doi.org/10.1111/1751-2980.12134 (2014).

Author contributions

Author Contributions: Conceptualization, M.M. (Miguel Mascarenhas), M.J.A. (Maria João Almeida) and M.M. (Miguel Martins); resources, J.A., T.R., P.C., J.M. ; writing – original draft preparation, M.M. (Miguel Mascarenhas), M.J.A. (Maria João Almeida) and M.M. (Miguel Martins); writing – review and editing, M.M. (Miguel Mascarenhas), M.J.A. (Maria João Almeida) and M.M. (Miguel Martins), T.R. and F.M.; supervision: M.G.H, B.A.C., J.W., A.C., Y.F., F.V.B., P.P. and G.M.; project administration - M.M. (Miguel Mascarenhas) and G.M. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Declarations

Competing interests

The authors declare no competing interests.

Ethics and consent to participate declarations

The study was performed following approval from the ethics committees of Centro Hospitalar Universitário de São João/Faculdade de Medicina da Universidade do Porto (CE 41/2021), Hospital Universitario Puerta de Hierro Majadahonda (PI 153/22) and New York University Langone Hospital (IRB 03845978/2023). This was a retrospective non interventional study conducted in accordance with the Declaration of Helsinki. Therefore, there was not a need to obtain written informed consent from each subject, after evaluation by the ethics committee of each institution.

Additional information

Correspondence and requests for materials should be addressed to M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.