OXFORD

Short Report

BJS, 2023, **110**, 1623–1627
https://doi.org/10.1093/bjs/znad201
Advance Access Publication Date: 4 July 2023
Short Report

Artificial intelligence-based models to assess the risk of malignancy on radiological imaging in patients with intraductal papillary mucinous neoplasm of the pancreas: scoping review

Alberto Balduzzi[1,†], Boris V. Janssen[2,3,4,†], Matteo De Pastena[1], Tommaso Pollini[1], Giovanni Marchegiani[1], Henk Marquering[4,5,6], Jaap Stoker[4,6], Inez Verpalen[4,6], Claudio Bassi[1], Marc G. Besselink[2,4,‡*] and Roberto Salvia[1,‡*]; for the Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium

[1]Department of Surgery and Oncology, Unit of General and Pancreatic Surgery, University of Verona Hospital Trust, Verona, Italy
[2]Department of Surgery, Amsterdam UMC, location University of Amsterdam, Amsterdam, the Netherlands
[3]Department of Pathology, Amsterdam UMC, location University of Amsterdam, Amsterdam, the Netherlands
[4]Cancer Centre Amsterdam, Amsterdam, the Netherlands
[5]Department of Biomedical Engineering and Physics, Amsterdam UMC, location University of Amsterdam, Amsterdam, the Netherlands
[6]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Location University of Amsterdam, Amsterdam, the Netherlands

Correspondence to: Marc G. Besselink, Department of Surgery, Cancer Centre Amsterdam, Amsterdam UMC, University of Amsterdam, De Boelelaan 1117 (VUMC hospital, ZH-7F18), 1081 HV Amsterdam, the Netherlands (e-mail: m.g.besselink@amsterdamumc.nl); Roberto Salvia, Unit of General and Pancreatic Surgery—Pancreas Institute Verona, Department of Surgery, Dentistry, Paediatrics and Gynaecology, University of Verona Hospital Trust, Policlinico G.B. Rossi, Piazzale L.A. Scuro 10, 37134 Verona, Italy (e-mail: roberto.salvia@univr.it)
[†]AB and BVJ share first authorship.
[‡]MGB and RS share senior authorship.

Introduction

Increased detection of pancreatic cystic neoplasms has drawn the attention of the medical community[1,2]. Among these, intraductal papillary mucinous neoplasms (IPMNs) represent a serious challenge for clinicians because of their (low) premalignant potential. Despite extensive efforts, the treatment of IPMN remains controversial, which is reflected by differences in the current three major guidelines[1,3,4].

Most patients diagnosed with IPMN will be kept under surveillance, aimed at monitoring progression of the cyst, which may require surgical resection in highly selected patients. Still, the risk of clinicians missing IPMN progression to malignancy is concerning[5], with burdensome consequences for the patient. This concern must be balanced against the risk of complications after major pancreatic surgery. Therefore, patient selection is crucial both to avoid unnecessary surgery for benign lesions, and to continue surveillance safely. Typically, diagnostic imaging plays a central role in guiding patient selection for, and the timing of, surgery. However, current imaging approaches fall short for optimal decision-making.

Machine learning assessment of radiological imaging may improve the assessment of IPMNs and add to the decision-making for surgery. This scoping review provides an overview of the available evidence on this topic.

Methods

Literature search

The Joanna Briggs Institute and PRISMA Extension for Scoping Reviews criteria were used for this scoping review[6,7]. The review was performed in PubMed and Embase up to 16 December 2021. The search strategy is provided in *Table S1*. Two assessors worked independently on literature screening, evaluation of eligibility, and inclusion, with conflicts handled through discussion. The remaining literature was subjected to full-text analysis. Original studies on imaging-based machine learning models (*Table S2*) in IPMN, which reported on model performance in terms of malignancy assessment, were included.

Data extraction and analysis

Two reviewers extracted data independently. If consensus could not be established, disagreements were resolved by discussion. When these two reviewers could not reach agreement, a third independent assessor was involved. Data from the included studies were analysed descriptively. The primary outcome was the discriminatory ability of the models measured by the area under the curve (AUC), accuracy, C-index, and *P* value. Model performance of 0.75 or more was considered sufficient for research reporting on AUC, C-index, and accuracy.

Specific attention was paid to determining whether the discriminative models were compared with the reference standard of clinical care based on guidelines.

Methodological assessment

The methodological quality of the included studies was assessed using the modified Radiomics Quality Score (mRQS)[8]. Within the mRQS (*Table S3*), radiomics approaches can reach a total of 36 points on 16 aspects, and deep learning approaches can reach a total of 32 points on 14 aspects. Two independent assessors

Received: February 21, 2023. **Revised:** April 13, 2023. **Accepted:** June 13, 2023

performed the methodological assessment. Discrepancies were resolved by discussion.

Results

Literature search

The literature search yielded 49 studies, of which 33 were excluded based on title and abstract screening, and four more studies after full-text screening. Eventually, 12 studies[9–20] fulfilled the eligibility criteria and were included in this scoping review (*Fig. S1*). *Table 1* details the included studies and the data extracted.

Radiomics models

Of the 10 IPMN radiomics models, 9 were based on CT and 1 on MRI. All models aimed to distinguish low- or intermediate-grade dysplasia from high-grade dysplasia or IPMN with concomitant pancreatic ductal adenocarcinoma. The studies had AUCs ranging between 0.76 and 0.96[9–17]; one[18] showed that two features were independent variables for malignant IPMN with ORs ranging from 1.49 to 1.52 and 0.977 to 0.981.

Deep learning models

Of the two deep learning models, one was based on MRI and the other on endoscopic ultrasound imaging. The models aimed to distinguish low- or intermediate-grade dysplasia from high-grade dysplasia or IPMN with concomitant pancreatic ductal adenocarcinoma. These models reached AUCs of 0.78 and 0.98[19,20].

Comparison with reference standard and added value

Two of the included studies compared the developed model with available guidelines. Corral *et al.*[20] reported a similar diagnostic performance for MRI-based deep learning (AUC 0.78), and American Gastroenterology Association[4] and Fukuoka (AUC 0.77)[1] guidelines. Conversely, the CT and MRI-based radiomics model of Cheng *et al.*[15] had superior discriminative performance (MRI: AUC 0.94; CT: AUC 0.86) to that of the clinical and imaging model based on the Fukuoka guidelines (AUC 0.77)[1].

Methodological assessment

The median mRQS score for the radiomics studies was 11.5 of 36, ranging from 4 to 17. The deep learning studies had a median

Table 1 Studies on imaging-based machine learning models assessing the malignant potential of intraductal papillary mucinous neoplasm

Reference	Country	Imaging	IPMN type	Target variable	Total *n*	Machine learning component	Data type	Outcome(s)
Hanania *et al.*[9]	USA	CT	BD and MD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 52 (10-fold cross-validation)	PCA and LR	Radiomics	AUC 0.96
Permuth *et al.*[10]	USA	CT	BD, MD, and Mixed	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 38 (10-fold cross-validation)	LR and PCA	Radiomics	AUC 0.93
Attiyeh *et al.*[11]	USA	CT	BD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 103 (10-fold cross-validation)	RF	Radiomics	AUC 0.76
Chakraborty *et al.*[12]	USA	CT	BD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 103 (10-fold cross-validation)	RF and SVM	Radiomics	AUC 0.77
Kuwahara *et al.*[19]	Japan	EUS	Undefined	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 50 (10-fold cross-validation)	NN	Image	AUC 0.98
Corral *et al.*[20]	USA	MRI	BD and MD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 139 (10-fold cross-validation)	NN	Image	AUC 0.78
Jeon *et al.*[18]	South Korea	MRI	BD, MD, and mixed	Low-grade IPMN *versus* high-grade IPMN or PDAC	248 (no validation)	LR	Radiomics	Entropy: OR 1.49–1.52 Compactness 2: OR 0.977–0.981
Harrington *et al.*[13]	USA	CT and EUS	BD and MD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Training 103 Testing 33	RF	Radiomics	AUC 0.83
Polk *et al.*[14]	USA	CT	Undefined	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 51 (5-fold cross-validation)	LR	Radiomics	AUC 0.90
Tobaly *et al.*[17]	France	CT	BD, MD, and Mixed	Low grade IPMN *versus* high grade IPMN or PDAC	Training 296 Testing 112	LASSO and LR	Radiomics	AUC 0.84
Cui *et al.*[16]	China	MRI	BD	Low-grade IPMN *versus* high-grade IPMN or PDAC	Training 107 Testing 99	LASSO and LR	Radiomics	AUC 0.81–0.82
Cheng *et al.*[15]	China	CT and MRI	Undefined	Low-grade IPMN *versus* high-grade IPMN or PDAC	Total 60 (10-fold cross-validation)	LR and SVM	Radiomics	MRI: AUC 0.879–0.940 CT: AUC 0.811–0.864

IPMN, intraductal papillary mucinous neoplasm; BD, branch duct; MD, main duct; PDAC, pancreatic ductal adenocarcinoma; PCA, principle component analysis; LR, logistic regression; AUC, area under the curve; EUS, endoscopic ultrasound imaging; RF, random forest; SVM, support vector machine; NN, neural network; LASSO, least absolute shrinkage and selection operator.

	1. Image protocol quality	2. Multiple segmentations	3. Phantom study on all scanners	4. Imaging at multiple time points	5. Feature reduction or adjustment for multiple testing	6. Multivariable analysis with non-imaging features	7. Detect and discuss biological correlates	8. Cut-off analyses	9. Discrimination statistics	10. Calibration statistics	11. Prospective study registered in a trial database	12. Validation	13. Comparison with standard	14. Potential clinical utility	15. Cost-effectiveness analysis	16. Open science and data	Total score
Hanania *et al.*[9]	0	0	0	0	–3	0	1	0	2	0	0	2	0	2	0	0	4 of 36
Permuth *et al.*[10]	1	1	0	0	3	1	1	0	2	0	0	2	0	2	0	0	13 of 36
Attiyeh *et al.*[11]	1	0	0	0	3	1	1	0	2	0	0	2	0	2	0	0	12 of 36
Chakraborty *et al.*[12]	1	0	0	0	3	0	1	0	2	0	0	2	0	2	0	0	11 of 36
Kuwahara *et al.*[19]	0	0	0	0	–	1	1	–	2	0	0	2	0	2	0	0	8 of 32
Corral *et al.*[20]	1	0	0	0	–	0	1	–	2	0	0	2	2	2	0	0	10 of 32
Jeon *et al.*[18]	1	0	0	0	3	1	1	0	1	0	0	–5	0	2	0	0	4 of 36
Harrington *et al.*[13]	1	0	0	0	3	1	1	0	1	0	0	2	0	2	0	0	11 of 36
Polk *et al.*[14]	1	0	0	0	–3	1	1	0	2	0	0	2	0	2	0	0	6 of 36
Tobaly *et al.*[17]	1	0	0	0	3	1	1	0	2	0	0	3	0	2	0	0	13 of 36
Cui *et al.*[16]	1	1	0	1	3	1	1	0	2	2	0	3	0	2	0	0	17 of 36
Cheng *et al.*[15]	1	1	0	0	3	0	1	0	2	0	0	2	2	2	0	0	14 of 36

Fig. 1 Methodological assessment of included studies using modified Radiomics Quality Score

Green, all points given; orange, not all points given; red, no points or negative points given; –, not applicable to this study owing to data type.

mRQS score of 9 of 32, ranging from 8 to 10. The included studies consistently scored poorly (25 per cent or fewer of the studies received all points) on 10 of 16 items in the mRQS: 2, 3, 4, 8, 10, 11, 12, 13, 15, and 16. Four items scored consistently positive (at least 75 per cent of the studies received all points): 5, 7, 9, and 14. Finally, two items scored moderately consistently (received all points in more than 25 per cent and less than 75 per cent of the studies): 1 and 6. *Figure 1* summarizes the mRQS scoring.

Discussion

This scoping review identified 12 artificial intelligence-based models to assess the risk of malignancy in patients with IPMN on radiological imaging. Although model performance was generally promising, the methodological quality of the studies was relatively poor. Furthermore, none of the models were applied in a prospective clinical setting or determined the added value compared with current guidelines or a clinical expert panel. If methodologically robust models are developed, and evaluated in a prospective setting, they may have the potential to enhance decision-making in finding the best time for surgery in patients diagnosed with IPMN.

This review has several limitations. First, owing to publication bias, models that were not discriminative may not have been submitted or accepted for publication. The model performance presented in this review may therefore be optimistic. Second, all studies are based on retrospective surgical series. However, most IPMNs are addressed to surgery after surveillance if cyst progression is observed. The selection bias originating from including only patients with surgically resected tumours makes the value of these models unclear in the unresected population. Third, the methods of the included studies varied extensively. Therefore, extracting generalizable results from this overview is difficult.

Future research should concentrate on developing methodologically sound, generalizable, and clinically validated models. Multiple methodological elements are frequently missed or ignored, as is evident from the mRQS scores of the research included. Once robust and generalizable models have been constructed, their performance and value should be validated in clinical settings. Currently available studies have focused on assessing the discriminative performance of machine learning models for malignant IPMNs. However, ideally, models would exclude the presence of malignancy with a high negative

predictive value and 'safely' advise surveillance in patients who would have been selected for surgical treatment according to current criteria. This would represent a true added value to current clinical practice.

This scoping review has provided evidence that 12 artificial intelligence-based machine learning models have sufficient capacity to evaluate the risk of malignancy in IPMN. However, the methodological quality of the included studies is inadequate, and the clinical value of the proposed models has not been proven. As a result, caution should be advised when interpreting these results, and the findings must be corroborated by additional high-quality studies. Future research should focus on developing rigorous models and investigating their usefulness in clinical practice to ensure that they are dependable tools for assessing the risk of malignancy in IPMN.

Funding

The authors have no funding to declare.

Acknowledgements

A.B. and B.V.J. are joint first authors, and M.G.B. and R.S. are joint last authors, of this article.

Author contributions

Alberto Balduzzi (Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration), Boris Janssen (Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration), Matteo De Pastena (Data Curation, Writing - Review & Editing), Tommaso Pollini (Data Curation, Writing - Review & Editing), Giovanni Marchegiani (Conceptualization, Writing - Review & Editing, Methodology, Supervision), Henk Marquering (Conceptualization, Writing - Review & Editing, Methodology, Supervision), Jaap Stoker (Conceptualization, Writing - Review & Editing, Methodology, Supervision), Inez Verpalen (Conceptualization, Writing - Review & Editing, Methodology, Supervision), Claudio Bassi (Conceptualization, Writing - Review & Editing, Methodology, Supervision), Marc Besselink (Conceptualization, Writing - Review & Editing, Methodology, Supervision), and Roberto Salvia (Conceptualization, Writing - Review & Editing, Methodology, Supervision).

Disclosure

The authors declare no conflict of interest.

Supplementary material

Supplementary material is available at *BJS* online.

Data availability

The raw data required to reproduce the scoping review findings were extracted from published articles.

References

1. Tanaka M, Fernández-del Castillo C, Kamisawa T, Jang JY, Levy P, Ohtsuka T *et al.* Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. *Pancreatology* 2017;**17**:738–753
2. Balduzzi A, Salvia R, Löhr M. Risk stratification tools for branch-duct intraductal papillary mucinous neoplasms of the pancreas. *United European Gastroenterol J* 2022;**10**:145–146
3. Del Chiaro M, Besselink MG, Scholten L, Bruno MJ, Cahen DL, Gress TM *et al.* European evidence-based guidelines on pancreatic cystic neoplasms. *Gut* 2018;**67**:789–804
4. Vege SS, Ziring B, Jain R, Moayyedi P, Adams MA, Dorn SD *et al.* American gastroenterological association institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. *Gastroenterology* 2015;**148**:819–822
5. Balduzzi A, Marchegiani G, Pollini T, Biancotto M, Caravati A, Stigliani E *et al.* Systematic review and meta-analysis of observational studies on BD-IPMNS progression to malignancy. *Pancreatology* 2021;**21**:1135–1145
6. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;**169**:467–473
7. Lockwood C, Tricco AC. Preparing scoping reviews for publication using methodological guides and reporting standards. *Nurs Health Sci* 2020;**22**:1–4
8. Janssen BV, Verhoef S, Wesdorp NJ, Huiskens J, de Boer OJ, Marquering H *et al.* Imaging-based machine-learning models to predict clinical outcomes and identify biomarkers in pancreatic cancer: a scoping review. *Ann Surg* 2022;**275**:560–567
9. Hanania AN, Bantis LE, Feng Z, Wang H, Tamm EP, Katz MH *et al.* Quantitative imaging to evaluate malignant potential of IPMNs. *Oncotarget* 2016;**7**:85 776–85 784
10. Permuth JB, Choi J, Balarunathan Y, Kim J, Chen DT, Chen L *et al.* Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. *Oncotarget* 2016;**7**:85 785
11. Attiyeh MA, Chakraborty J, Gazit L, Langdon-Embry L, Gonen M, Balachandran VP *et al.* Preoperative risk prediction for intraductal papillary mucinous neoplasms by quantitative CT image analysis. *HPB (Oxford)* 2019;**21**:212–218
12. Chakraborty J, Midya A, Gazit L, Attiyeh M, Langdon-Embry L, Allen PJ *et al.* CT radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas. *Med Phys* 2018;**45**:5019–5029
13. Harrington KA, Williams TL, Lawrence SA, Chakraborty J, Al Efishat MA, Attiyeh MA *et al.* Multimodal radiomics and cyst fluid inflammatory markers model to predict preoperative risk in intraductal papillary mucinous neoplasms. *J Med Imaging (Bellingham)* 2020;**7**:1
14. Polk SL, Choi JW, McGettigan MJ, Rose T, Ahmed A, Kim J *et al.* Multiphase computed tomography radiomics of pancreatic intraductal papillary mucinous neoplasms to predict malignancy. *World J Gastroenterol* 2020;**26**:3458–3471
15. Cheng S, Shi H, Lu M, Wang C, Duan S, Xu Q *et al.* Radiomics analysis for predicting malignant potential of intraductal papillary mucinous neoplasms of the pancreas: comparison of CT and MRI. *Acad Radiol* 2022;**29**:367–375
16. Cui S, Tang T, Su Q, Wang Y, Shu Z, Yang W *et al.* Radiomic nomogram based on MRI to predict grade of branching type intraductal papillary mucinous neoplasms of the pancreas: a multicenter study. *Cancer Imaging* 2021;**21**:26
17. Tobaly D, Santinha J, Sartoris R, Burgio MD, Matos C, Cros J *et al.* CT-based radiomics analysis to predict malignancy in patients with intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Cancers (Basel)* 2020;**12**:3089

18. Jeon SK, Kim JH, Yoo J, Kim JE, Park SJ, Han JK. Assessment of malignant potential in intraductal papillary mucinous neoplasms of the pancreas using MR findings and texture analysis. *Eur Radiol* 2021;**31**:3394–3404
19. Kuwahara T, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M *et al.* Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas. *Clin Transl Gastroenterol* 2019;**10**:1–8
20. Corral JE, Hussein S, Kandel P, Bolan CW, Bagci U, Wallace MB. Deep learning to classify intraductal papillary mucinous neoplasms using magnetic resonance imaging. *Pancreas* 2019; **48**:805–810