— Review Article—

Pooled diagnostic parameters of artificial intelligence in EUS image analysis of the pancreas: A descriptive quantitative review

Babu P. Mohan[1], Antonio Facciorusso[2], Shahab R. Khan[3], Deepak Madhu[4], Lena L. Kassab[5], Suresh Ponnada[6], Saurabh Chandan[7], Stefano F. Crino[8], Gursimran S. Kochhar[9], Douglas G. Adler[1], Michael B. Wallace[10]

[1]Department of Gastroenterology and Hepatology, University of Utah, Salt Lake City, Utah, USA; [2]Gastroenterology Unit, University of Foggia, Foggia, Italy; [3]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [4]Department of Gastroenterology, Aster MIMS, Calicut, Kerala, India; [5]Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA; [6]Department of Internal Medicine, Roanoke Medical Center, Roanoke, Virginia, USA; [7]Department of Gastroenterology, CHI Creighton University Medical Center, Omaha, NE, USA; [8]Gastroenterology and Digestive Endoscopy Unit, The Pancreas Institute, G. B. Rossi University Hospital, Verona, Italy; [9]Department of Gastroenterology and Hepatology, Allegheny Health Network, Pittsburgh, PA, USA; [10]Department of Gastroenterology and Hepatology, Sheikh Shahkbout Medical City, Abu Dhabi, UAE

ABSTRACT

EUS is an important diagnostic tool in pancreatic lesions. Performance of single-center and/or single study artificial intelligence (AI) in the analysis of EUS-images of pancreatic lesions has been reported. The aim of this study was to quantitatively study the pooled rates of diagnostic performance of AI in EUS image analysis of pancreas using rigorous systematic review and meta-analysis methodology. Multiple databases were searched (from inception to December 2020) and studies that reported on the performance of AI in EUS analysis of pancreatic adenocarcinoma were selected. The random-effects model was used to calculate the pooled rates. In cases where multiple 2 × 2 contingency tables were provided for different thresholds, we assumed the data tables as independent from each other. Heterogeneity was assessed by I^2% and 95% prediction intervals. Eleven studies were analyzed. The pooled overall accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were 86% (95% confidence interval [82.8–88.6]), 90.4% (88.1–92.3), 84% (79.3–87.8), 90.2% (87.4–92.3) and 89.8% (86–92.7), respectively. On subgroup analysis, the corresponding pooled parameters in studies that used neural networks were 85.5% (80–89.8), 91.8% (87.8–94.6), 84.6% (73–91.7), 87.4% (82–91.3), and 91.4% (83.7–95.6)], respectively. Based on our meta-analysis, AI seems to perform well in the EUS-image analysis of pancreatic lesions.

Key words: artificial intelligence, endoscopic ultrasound, meta-analysis

Access this article online

Quick Response Code:

Website: www.eusjournal.com

DOI: 10.4103/EUS-D-21-00063

How to cite this article: Mohan BP, Facciorusso A, Khan SR, Madhu D, Kassab LL, Ponnada S, *et al*. Pooled diagnostic parameters of artificial intelligence in EUS image analysis of the pancreas: A descriptive quantitative review. Endosc Ultrasound 2022;11:156-69.

Address for correspondence

Dr. Babu P. Mohan, Department of Gastroenterology and Hepatology, University of Utah Health, 30 N 1900 E, SOM 4R118, Salt Lake City, UT 84132, USA. E-mail: dr.babu.pm@gmail.com

Received: 2021-03-04; **Accepted:** 2021-11-29; **Published online:** 2022-03-21

INTRODUCTION

EUS has become an indispensable investigation tool in the disorders of the pancreas.[1] EUS-guided sampling, by means of fine-needle aspiration (FNA) and/or fine-needle biopsy (FNB), have demonstrated sensitivity rates ranging from 74% to 95% in the diagnosis of pancreatic malignancy.[1,2] However, the diagnosis of solid pancreatic lesions continues to be a challenge, especially in the presence of background chronic pancreatitis.[1,3] Clinical decision-making can be difficult when tissue sampling is negative and/or inconclusive. In such circumstances, the physician cannot conclude the lesion to be benign if there is a high degree of clinical suspicion of malignancy, due to the extremely poor prognosis associated with pancreatic malignancy.[4]

The reported sensitivity of EUS is 50%–60% in the diagnosis of solid lesions of the pancreas.[1,3] Circumstances arise when EUS by itself is not an adequate tool. To help improve the diagnostic performance, EUS-image enhancement with the aid of contrast-enhanced EUS and techniques such as EUS-elastography have been introduced. The reported accuracy of diagnosing pancreatic tumors with the addition of these modalities is about 80%–90%.[1-3,5,6]

The exceptional performance of AI in medical diagnosis using deep learning algorithm in computer vision is creating a new hype, as well as hope. Recently, data have emerged on the use of artificial intelligence (AI) in computer-aided diagnosis of lesions seen on endoscopic images and multiple studies have summarized their pooled performances.[7-10] Similarly, recent evidence has emerged on the utility of AI in the analysis of EUS images of pancreatic lesions.[11,12] However, the data is currently evolving and limited.[5,13-23]

We conducted this systematic review and meta-analysis to consolidate and appraise the reported literature on the use of AI in EUS evaluation of solid lesions of the pancreas. Due to the evolving nature of the topic, we expected potential variability in terms of the clinical situation, and machine learning algorithms that might contribute to considerable heterogeneity. In this study, we aim to present descriptive pooled estimates rather than precise point estimates.

METHODS

Search strategy

A medical librarian searched the literature for the concepts of AI in EUS analysis of pancreatic disorders. The search strategies were created using a combination of keywords and standardized index terms. Searches were run in December 2020 in ClinicalTrials.gov, Ovid EBM Reviews, Ovid, Embase (1974+), Ovid Medline (1946 + including Epub ahead of print, in-process and other nonindexed citations), Scopus (1970+) and Web of Science (1975+). Results were limited to the English language. All results were exported to Endnote X9 (Clarivate Analytics) where obvious duplicates were removed leaving 4245 citations. The search strategy is provided in Appendix 1. The MOOSE checklist was followed and is provided as Appendix 2.[24] Reference lists of evaluated studies were examined to identify other studies of interest.

Study selection

In this meta-analysis, we included studies that tested AI learning models for the detection and/or differentiation of pancreatic mass lesions on EUS. Studies were included irrespective of the machine learning algorithm, inpatient/outpatient setting; study sample-size, follow-up time, abstract/manuscript status, and geography as long as they provided the appropriate data needed for the analysis.

Our exclusion criteria were as follows: (1) nonclinical studies that reported on the mathematical development of convolutional neural network (CNN) algorithms, and (2) studies not published in the English language. In cases of multiple publications from a single research group reporting on the same patient cohort and/or overlapping cohorts, each reported contingency tables were treated as being mutually exclusive. When needed, authors were contacted via E-mail for clarification of data and/or study-cohort overlap.

Data abstraction and definitions

Data on study-related outcomes from the individual studies were abstracted independently onto a predefined standardized form by at least two authors (BPM, SRK). Disagreements were resolved by consultation with another author (AF). Diagnostic performance data was extracted, and contingency tables were created at the reported thresholds. Contingency tables consisted of reported accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive

value (NPV). If a study provided multiple contingency tables for the same or for different algorithms, we assumed these to be independent from each other. This assumption was accepted, as the goal of the study was to provide an overview of the pooled rates of various studies rather than providing precise point estimates.

Assessment of study quality

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to assess the quality and potential bias of all studies by two authors independently (BPM, DM).[25] Conflicts were resolved with discussion and involvement of a third author (SC). Four domains, namely patient selection, index test, reference standard, flow, and timing, were assessed. Two categories: Risk of bias and applicability concerns were assessed under the domains of patient selection, index test, and reference standard. The risk of bias was also assessed in the domain of flow and timing.

Statistical analysis

We used meta-analysis techniques to calculate the pooled estimates in each case following the random-effects model.[26] We assessed heterogeneity between study-specific estimates by using Cochran *Q* statistical test for heterogeneity, 95% prediction interval (PI), which deals with the dispersion of the effects, and the I^2 statistics.[27,28] A formal publication bias assessment was not planned due to the nature of the pooled results being derived from the studies. All analyses were performed using Comprehensive Meta-Analysis (CMA) software, version 3 (BioStat, Englewood, NJ).

RESULTS

Search results and study characteristics

The literature search resulted in 4245 study hits [Figure 1]. All 4245 studies were screened and 39 full-length articles and/or abstracts were assessed that reported on the performance of AI in EUS. After removing irrelevant articles, eleven studies were included in the final analysis.[5,13,15-23] Study by Kuwahara *et al.*, assessed the ability of AI to predict malignancy in IPMN lesions and therefore was not included.[14] The study selection flow chart is illustrated in Figure 1.

Based on the revised QUADAS-2 study assessment, unclear risk was noted with patient selection and flow and timing. Detailed assessment is illustrated in Supplementary Table 1. Seven studies evaluated the performance of AI on EUS images,[13,15-17,21-23] three on EUS elastography[5,19,20] and one on contrast-enhanced harmonic EUS.[18] Majority of the studies evaluated the use of AI in help detecting and/or differentiating pancreatic malignancy from chronic pancreatitis.[13,16,17,19,21-23] Whereas the study by Marya *et al.*, analyzed the ability of AI to diagnose autoimmune pancreatitis and we extracted data that reported on the performance of AI in pancreatic adenocarcinoma.[15] From the included studies, we were able to extract a total of 10 datasets for accuracy, 13 datasets each for sensitivity and specificity, 12 datasets each for PPV and NPV. The studies used a composite of pathological evaluation and expert evaluation of the images as the reference standard.

Meta-analysis outcomes

The pooled accuracy was 86% (95% confidence interval [CI] 82.8–88.6, I^2 = 57%) [Figure 2], sensitivity was 90.4% (95% CI 88.1–92.3, I^2 = 39%) [Figure 3], specificity was 84% (95% CI 79.3–87.8, I^2 = 88%) [Figure 4], positive predictive value was 90.2% (95% CI 87.4–92.3, I^2 = 70%) [Supplementary Figure 1] and negative predictive value was 89.8% (95% CI 86–92.7, I^2 = 90%) [Supplementary Figure 2].

In subgroup analysis of studies that exclusively used neural networks as the machine learning algorithm, the pooled accuracy was 85.5% (95% CI 80–89.8, I^2 = 69%) [Supplementary Figure 3], sensitivity was 91.8% (95% CI 87.8–94.6, I^2 = 45%) [Supplementary Figure 4], the specificity was 84.6% (95% CI 73.9–91.7, I^2 = 90%), [Supplementary Figure 5] the positive predictive value was 87.4% (95% CI 82–91.3, I^2 = 68%) [Supplementary Figure 6], and the negative predictive value was 91.4% (95% CI 83.7–95.6, I^2 = 85%) [Supplementary Figure 7]. Pooled rates are summarized in Table 2, along with the subgroup analysis based on analysis of EUS-images and EUS-elastography.

VALIDATION OF META-ANALYSIS RESULTS

Sensitivity analysis

To assess whether anyone study had a dominant effect on the meta-analysis, we excluded one study at a time and analyzed its effect on the main summary estimate. On this analysis, no single study significantly affected the outcome or the heterogeneity.

Table 1. Study characteristics

Study, year	Design, time period, center, country	Study aim	Image type	Machine learning model	Total images
Carrara, 2018	Prospective, December 2015-February 2017, Single-center, Italy	Characterization of solitary pancreatic lesions	EUS elastography	Fractal-based quantitative analysis	NR
Das, 2008	Retrospective, Single center, USA	Differentiate pancreatic adenocarcinoma from nonneoplastic tissue	EUS images	Neural network	11,099
Marya, 2020	Retrospective, Single center, USA	Data on pancreatic adenocarcinoma	EUS images/ videos	Neural network	1,174,461 (EUS images), 955 (EUS frames per second) (video data)
Norton, 2001	Retrospective, single center, USA	Differentiate malignancy from pancreatitis	EUS images	Neural network	NR
Ozkan, 2016	Retrospecitve, January 2013-September 2014, Single center, Turkey	Diagnosing pancreatic cancer	EUS images	Neural network	332 (202 cancer and 130 noncancer)
Saftoiu, 2008	Prospective, cross-sectional, multicenter, August 2005-November 2006 (Denmark), December 2006-September 2007 (Romania)	Differentiate malignancy from pancreatitis	EUS elastography	Neural network	NR
Saftoiu, 2012	Prospective, blinded, multicenter (13), Romania, Denmark, Germany, Spain, Italy, France, Norway, and United Kingdom	Diagnosis of focal pancreatic lesions	EUS elastography	Neural network	774
Saftoiu, 2015	Prospective, observational trial, multicenter (5), Romania, Denmark, Germany, and Spain	Diagnosis of focal pancreatic masses	CEH-EUS	Neural network	NR
Tonozuka, 2020	Prospective, April 2016-August 2019, Single center, Japan	Diagnosing pancreatic cancer	EUS images	Neural network	920 (endosonographic images), 470 (images were independently tested)
Zhang, 2010	Retrospective, Controlled, March 2005 and December 2007, Single center, China	Diagnosing pancreatic cancer	EUS images	SVM	NR
Zhu, 2013	Retrospective, May 2002-August 2011, Single center, China	Differentiate malignancy from pancreatitis	EUS images	SVM	NR

Study, year	Total patients	Accuracy	Sensitivity	Specificity	PPV	NPV
Carrara, 2018	100	85.3 (95% CI, 78.4-92.2) (pSR)/84.3 (95% CI, 76.5-91.2) (wSR)/84.31 (95% CI, 76.47-90.20) (both)	88.4 (95% CI, 79.7-95.7) (pSR)/91.3 (95% CI, 84.2-97.1) (wSR)/86.96 (95% CI, 78.26-94.20) (both)	78.8 (95% CI, 63.6-91.0) (pSR)/69.7 (95% CI, 54.6-84.9) (wSR)/78.79 (95% CI, 63.64-90.91) (both)	89.7 (95% CI, 83.5-95.5) (pSR)/86.5 (95% CI, 80.3-92.8) (wSR)/89.71 (95% CI, 83.10-95.38) (both)	76.9 (95% CI, 65.0-88.9) (pSR)/80.0 (95% CI, 66.7-92.6) (wSR)/74.29 (95% CI, 62.86-86.67) (both)
Das, 2008	56 (22 *n*; Group I [normal pancreas], 12 *n*; Group II [Chronic pancreatitis], 22 *n*; Group III [pancreatic adenocarcinoma])	100%	93% (95% CI, 89%-97%)	92% (95% CI, 88%-96%)	87% (95% CI, 82%-92%)	96% (95% CI, 93%-99%)
Marya, 2020	583	NR	0.95 (0.91-0.98)	0.91 (0.86-0.94)	0.87 (0.82-0.91)	0.97 (0.93-0.98)
Norton, 2001	35 (14 *n* [chronic pancreatitis], 21 *n* [pancreatic adenocarcinoma])	80%	100%	50%	75%	100%
Ozkan, 2016	172	87.50%	83.30%	93.30%	NR	NR
Saftoiu, 2008	68 (22 *n*=Normal pancrease), (11 *n*=Chronic pancreatitis), (32 *n*=Pancreatic adenocarcinoma), and (3 *n*=Pancreatic neuroendocrine tumors)	89.70%	91.40%	87.90%	88.90%	90.60%

Contd...

Table 1. Contd...

Study, year	Total patients	Accuracy	Sensitivity	Specificity	PPV	NPV
Saftoiu, 2012	258	84.27% (95% CI, 83.09%-85.44%)	87.59%	82.94%	96.25%	57.22%
Saftoiu, 2015	167 (112 *n*=Pancreatic carcinoma and 55 *n*=Chronic pancreatitis)	NR	94.64% (95% CI, 88.22%-97.80%)	94.44% (95% CI, 83.93%-98.58%)	97.24% (95% CI, 91.57%-99.28%)	89.47% (95% CI, 78.165-95.72%)
Tonozuka, 2020	139 (76 *n*=Pancreatic ductal carcinoma, 34 *n*=Chronic pancreatitis, and 29 *n*=Normal pancreas)	NR	92.40%	84.10%	86.80%	90.70%
Zhang, 2010	216 (153 *n* pancreatic cancer and 63 *n* [20 *n* normal pancreas and/or 43 *n* chronic pancreatitis] noncancer patients)	97.98% (1.23%)	94.32% (0.03%)	99.45% (0.01%)	98.65% (0.02%)	97.77% (0.01%)
Zhu, 2013	388 (262 *n*=Pancreatic carcinoma and 126 *n*=Chronic pancreatitis)	94.20% (0.1749%)	96.25% (0.4460%)	93.38% (0.2076%)	92.21% (0.4249%)	96.68% (0.1471%)

CEH: Contrast enhanced harmonic; SVM: Support vector machine; NR: Not reported; pSR: Parenchymal strain ratio; wSR: Wall strain ratio; PPV: Positive predictive value; NPV: Negative predictive value

Heterogeneity

We expected a large degree of between-study heterogeneity due to the broad nature of machine learning algorithms, EUS modalities, and varying diagnosis of pancreatic lesions included in this study. On subgroup analysis, the pooled rates of EUS elastography and pooled rates of studies that used neural network-based machine learning algorithms were noted be lower than the overall heterogeneity [Table 2].

Table 2. Summary of pooled rates

	Pooled rate (95% CI)	I^2% heterogeneity (95% PI)
Accuracy		
Overall	86% (82.8-88.6) 10 datasets	57% (71-94)
EUS-images	91.8% (82.3-96.4) 5 datasets	78% (52-99)
EUS-elastography	85.4% (82-88.2) 5 datasets	0% (79-89)
Neural network algorithm	85.5% (80-89.8) 5 datasets	69% (61-97)
Sensitivity		
Overall	90.4% (88.1-92.3) 13 datasets	39% (83-96)
EUS-images	93.4% (88.9-96.1) 7 datasets	60% (78-98)
EUS-elastography	88.9% (85.8-91.4) 5 datasets	0% (84-93)
Neural network algorithm	91.8% (87.8-94.6) 8 datasets	45% (84-97)
Specificity		
Overall	84% (79.3-87.8) 13 datasets	88% (51-97)
EUS-images	89.8% (76.3-96) 7 datasets	92% (35-99)
EUS-elastography	79.9% (73.5-85.1) 5 datasets	61% (55-93)
Neural network algorithm	84.6% (73-91.7) 8 datasets	90% (39-97)
PPV		
Overall	90.2% (87.4-92.3) 12 datasets	70% (65-97)
EUS-images	87.9% (80.8-92.6) 6 datasets	75% (54-96)
EUS-elastography	90% (86.6-92.6) 5 datasets	16% (85-95)
Neural network algorithm	87.4% (82-91.3) 7 datasets	68% (59-96)
NPV		
Overall	89.8% (86-92.7) 12 datasets	90% (51-99)
EUS-images	96.3% (93.3-98) 6 datasets	37% (89-98)
EUS-elastography	77% (65.1-85.8) 5 datasets	86% (27-96)
Neural network algorithm	91.4% (83.7-95.6) 7 datasets	85% (43-98)

CI: Confidence interval; PPV: Positive predictive value; NPV: Negative predictive value; PI: Prediction interval

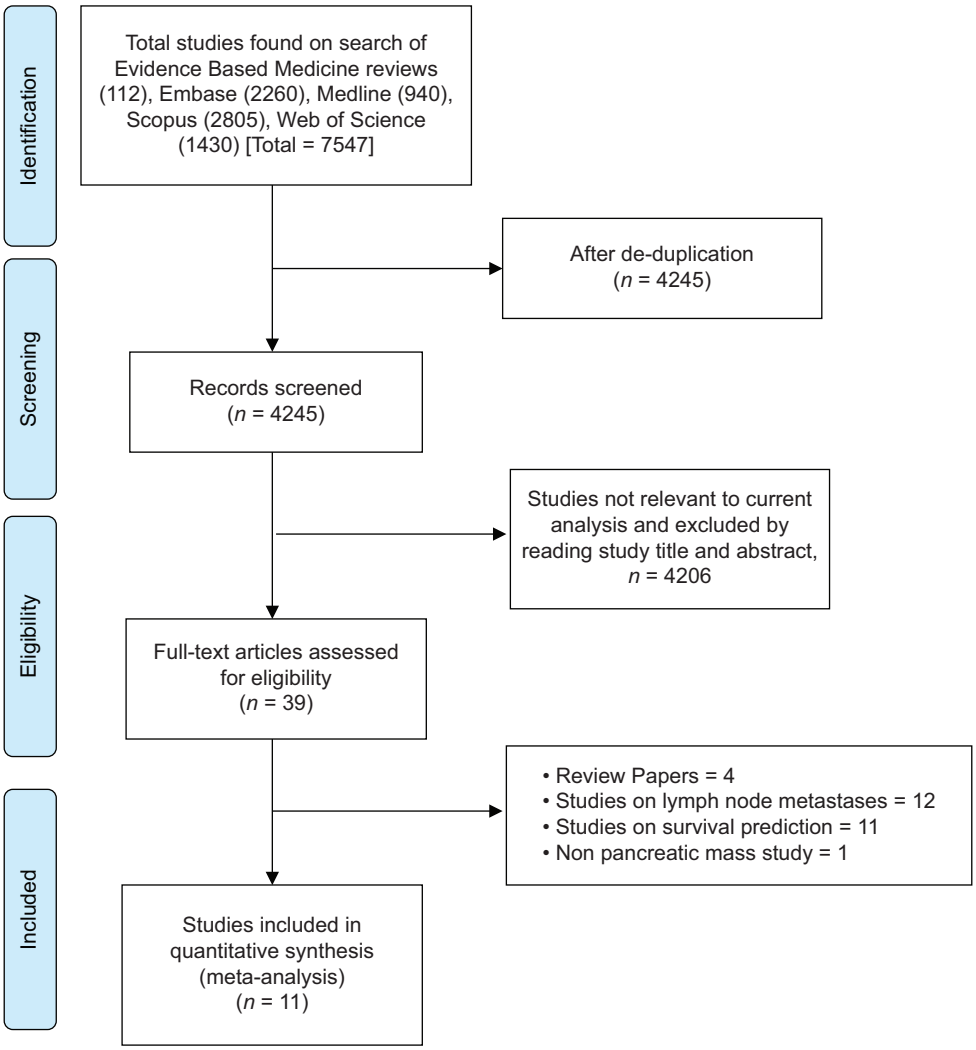


Figure 1. Study selection flow chart

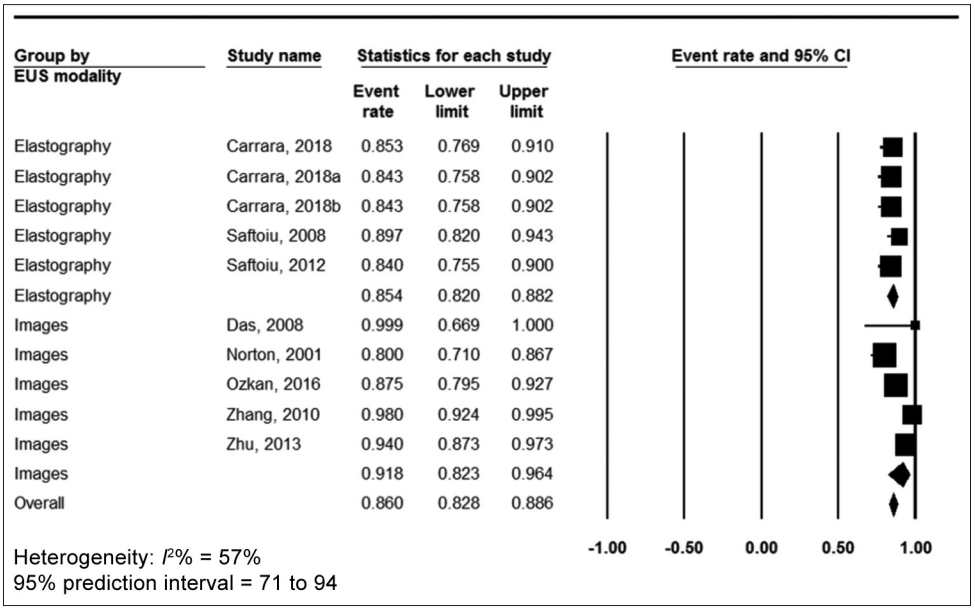


Figure 2. Forest plot, accuracy

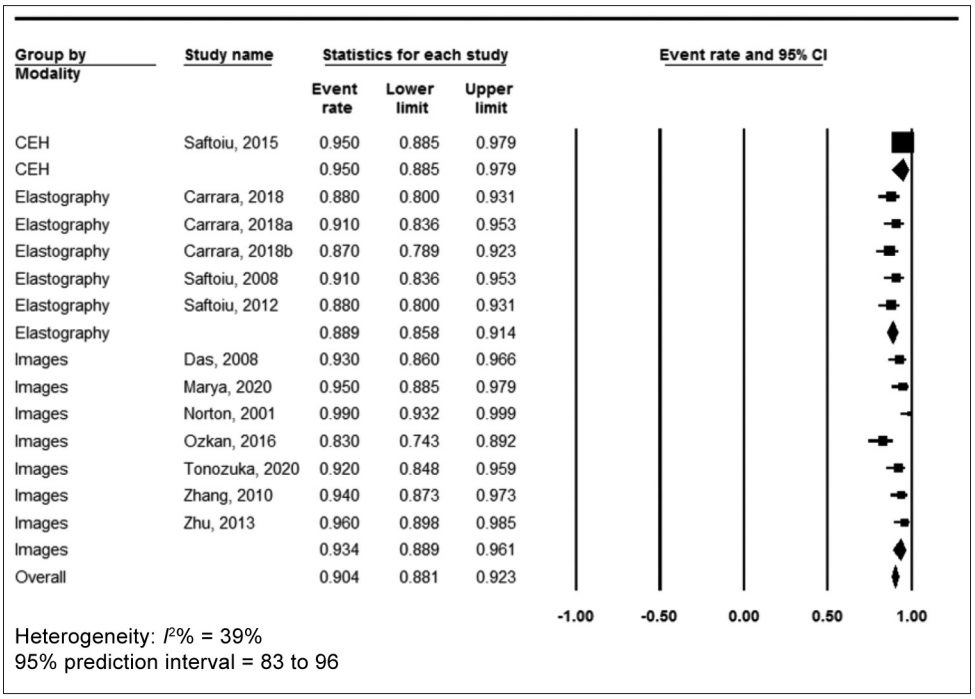


Figure 3. Forest plot, sensitivity

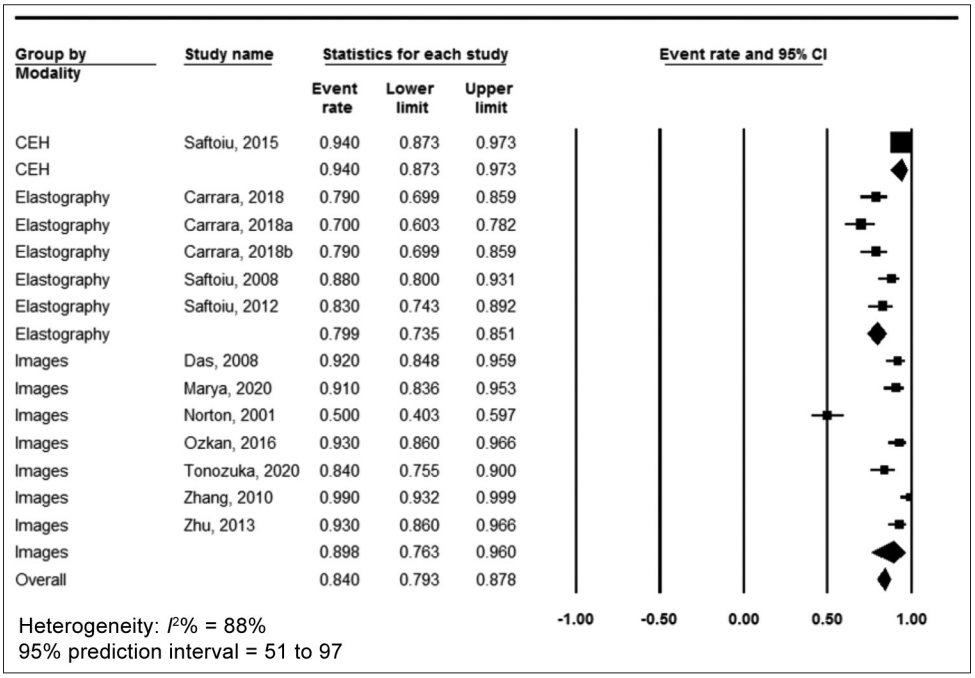


Figure 4. Forest plot, specificity

Publication bias

Publication bias assessment largely depends on the sample size and the reported effect size. A publication bias assessment was deferred in this study because the studied modality was AI and the reported effects were diagnostic parameters, both of which do not conform to the basics of publication bias assessment.[29]

DISCUSSION

In this systematic review and meta-analysis assessing AI-based machine learning in the assessment of pancreatic lesions on EUS imaging, we found that AI demonstrated a pooled accuracy of 86%, sensitivity of 90.4%, specificity of 84%, PPV of 90.2%, and NPV of 89.8%, albeit with expected heterogeneity.

EUS is not always able to differentiate neoplasia from reactive changes, especially in the presence of chronic pancreatitis. Pancreatic cancer is one of the most heterogeneous neoplastic diseases, owing to the complex nature of tissue and cell groups within the organ that is complicated by the extensively dense fibroblastic stroma and blood flow variations. In addition, there exists extensive spectrum of molecular subtypes determined by a variable number of gene mutations. Furthermore, the yield of EUS-guided FNA and/or FNB is heavily dependent on accurate targeting of the area of interest-based on the interpretation of the EUS images. Can AI prove to be a helpful computer aid to the therapeutic endoscopist in this regard?

Although premature for clinical application, this study demonstrates the high diagnostic performance of AI in the interpretation of lesions of the pancreas based on EUS images. We report an overall pooled NPV of 89.8% that is pretty close to the threshold proposed by The American Society of Gastrointestinal Endoscopy Preservation Incorporation of Valuable Endoscopic Innovations-2 of 90% or greater for real-time optical diagnosis using advanced endoscopic imaging.[30] This target was achieved in the subgroup analysis of the assessment of EUS-images (NPV = 96.3%) and in studies that exclusively used neural networks as the machine learning algorithm (NPV = 91.4%).

How do these results compare to the current practice of EUS-FNA and/or FNB? Although we did not have direct comparison cohorts, we can put the results of this study in perspective to the currently reported data in the literature. Based on meta-analyses data, the pooled sensitivity and specificity of EUS-FNA in the diagnosis of pancreatic cancer are 85%–89% and 96%–98%, respectively.[31,32] Comparable results have been reported with EUS-guided FNB of pancreatic masses, and moreover, EUS-FNB with newer EUS specific core-biopsy needles like Franseen and Fork-Tip needles have demonstrated superior accuracy rates.[33-37] Based on the results of this study, one can hypothesize superior diagnostic results with the combination of AI and newer core-biopsy needles in the EUS evaluation of solid pancreatic lesions.

The type of machine learning algorithm developed is important and deep learning by means of CNN has been shown to be exceptionally superior when compared to other algorithms in the computer-vision-based analysis of images.[38] CNNs are able to process data in various forms and of particular interest to the medical field is the image and video-based learning. The architecture of CNN is designed as a series of layers, particularly convolutional and pooling layers, followed by fully connected layers.[38] The important prerequisite for a high-performing algorithm is huge amounts of training data. Based on this analysis, neural network-based analysis of EUS in lesions of the pancreas demonstrated an accuracy of 85.5%, sensitivity of 91.8%, specificity of 84.6%, PPV of 87.4%, and NPV of 91.4%. In the recently published study by Tonozuka *et al.*, authors used a CNN to train EUS-images in the detection of pancreatic cancer and reported high diagnostic parameters that were comparable to a human's ability of image recognition.[21]

Although, an AI-based computer-aid seems promising in the analysis of EUS images of pancreatic lesions, current data needs to be interpreted with caution and the following limitations of machine learning need to be acknowledged. The included studies evaluated the performance of AI in experimental conditions. Prospective real-life scenario studies do not exist at this time. There was the lack in uniformity of validating the training process of the algorithm before using it for testing. Moreover, studies varied between differentiation of pancreatic malignancy from chronic pancreatitis and detection of lesions of EUS. In the near future, we can expect further studies exploring deep learning algorithms by means of CNN in EUS-image analysis of pancreatic lesions. To enable robust training of such algorithms, a global, open-source, correctly labeled EUS-image repository akin to Google-ImageNet should be explored.

We acknowledge that the data were heterogeneous. However, the high heterogeneity should not be considered of a major issue here as it is well-known that I^2% statistics is higher when considering continuous variables as compared to categorical outcomes due to the intrinsic numeric nature of these variables.[39] Therefore, I^2% values should be interpreted with caution here and moreover, in a proportion meta-analysis like ours, heterogeneity does not reflect a different direction in the pooled effects. Nevertheless, this study demonstrates descriptive pooled estimates of diagnostic parameters achievable by well-conducted studies in future, and variables such as the EUS modality, machine learning algorithm, and underlying disease should be kept consistent as much as possible.

CONCLUSIONS

Based on our analysis, AI seemed to perform well in the analysis of EUS images of pancreatic lesions. The prerequisites are to achieve high sensitivity and NPV, which our study demonstrates, however real-life clinical scenario studies are warranted to establish the role of AI in daily EUS practice of analyzing the pancreas.

Acknowledgements

Dana Gerberi, MLIS, Librarian, Mayo Clinic Libraries, for help with the systematic literature search.

Unnikrishnan M. Pattath, BTech, MBA; Data Science Architect (India) for help with technical details on convolutional neural networks and other machine learning algorithms.

Supplementary materials

Supplementary information is linked to the online version of the paper on the Endoscopic Ultrasound website.

Financial support and sponsorship

Nil.

Conflicts of interest

Douglas G. Adler is a Co-Editor-in-Chief of the journal. This article was subject to the journal's standard procedures, with peer review handled independently of this editor and his research groups.

REFERENCES

1. Iglesias-Garcia J, Lindkvist B, Lariño-Noia J, *et al*. Endoscopic ultrasound elastography. *Endosc Ultrasound* 2012;1:8-16.
2. Maguchi H. The roles of endoscopic ultrasonography in the diagnosis of pancreatic tumors. *J Hepatobiliary Pancreat Surg* 2004;11:1-3.
3. Iglesias-García J, Lariño-Noia J, Lindkvist B, *et al*. Endoscopic ultrasound in the diagnosis of chronic pancreatitis. *Rev Esp Enferm Dig* 2015;107:221-8.
4. McGuigan A, Kelly P, Turkington RC, *et al*. Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. *World J Gastroenterol* 2018;24:4846-61.
5. Carrara S, Di Leo M, Grizzi F, *et al*. EUS elastography (strain ratio) and fractal-based quantitative analysis for the diagnosis of solid pancreatic lesions. *Gastrointest Endosc* 2018;87:1464-73.
6. Sakamoto H, Kitano M, Suetomi Y, *et al*. Utility of contrast-enhanced endoscopic ultrasonography for diagnosis of small pancreatic carcinomas. *Ultrasound Med Biol* 2008;34:525-32.
7. Goodfellow I, Bengio Y, Courville A, *et al*. Deep Learning. Cambridge: MIT Press; 2016.
8. Mohan BP, Facciorusso A, Khan SR, *et al*. Real-time computer aided colonoscopy versus standard colonoscopy for improving adenoma detection rate: A meta-analysis of randomized-controlled trials. *EClinicalMedicine* 2020;29-30:100622.
9. Mohan BP, Khan SR, Kassab LL, *et al*. High pooled performance of convolutional neural networks in computer-aided diagnosis of GI ulcers and/or hemorrhage on wireless capsule endoscopy images: A systematic review and meta-analysis. *Gastrointest Endosc* 2021;93:356-64.e4.
10. Mohan BP, Khan SR, Kassab LL, *et al*. Accuracy of convolutional neural network-based artificial intelligence in diagnosis of gastrointestinal lesions based on endoscopic images: A systematic review and meta-analysis. *Endosc Int Open* 2020;8:E1584-94.
11. Kuwahara T, Hara K, Mizuno N, *et al*. Current status of artificial intelligence analysis for endoscopic ultrasonography. *Dig Endosc* 2021;33:298-305.
12. Tonozuka R, Mukai S, Itoi T. The role of artificial intelligence in endoscopic ultrasound for pancreatic disorders. *Diagnostics (Basel)* 2020;11:18.
13. Das A, Nguyen CC, Li F, *et al*. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. *Gastrointest Endosc* 2008;67:861-7.
14. Kuwahara T, Hara K, Mizuno N, *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.
15. Marya NB, Powers PD, Chari ST, *et al*. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* 2021;70:1335-44.
16. Norton ID, Zheng Y, Wiersema MS, *et al*. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. *Gastrointest Endosc* 2001;54:625-9.
17. Ozkan M, Cakiroglu M, Kocaman O, *et al*. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. *Endosc Ultrasound* 2016;5:101-7.
18. Săftoiu A, Vilmann P, Dietrich CF, *et al*. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). *Gastrointest Endosc* 2015;82:59-69.
19. Săftoiu A, Vilmann P, Gorunescu F, *et al*. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. *Gastrointest Endosc* 2008;68:1086-94.
20. Săftoiu A, Vilmann P, Gorunescu F, *et al*. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol* 2012;10:84-90.e1.
21. Tonozuka R, Itoi T, Nagata N, *et al*. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: A pilot study. *J Hepatobiliary Pancreat Sci* 2021;28:95-104.
22. Zhang MM, Yang H, Jin ZD, *et al*. Differential diagnosis of pancreatic cancer from normal tissue with digital imaging processing and pattern recognition based on a support vector machine of EUS images. *Gastrointest Endosc* 2010;72:978-85.
23. Zhu M, Xu C, Yu J, *et al*. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: A diagnostic test. *PLoS One* 2013;8:e63820.
24. Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.
25. Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
26. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
27. Mohan BP, Adler DG. Heterogeneity in systematic review and meta-analysis: How to read between the numbers. *Gastrointest Endosc* 2019;89:902-3.
28. Higgins JP, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
29. Easterbrook PJ, Berlin JA, Gopalan R, *et al*. Publication bias in clinical research. *Lancet* 1991;337:867-72.
30. ASGE Technology Committee; Abu Dayyeh BK, Thosani N, Konda V, *et al*. ASGE Technology Committee systematic review and meta-analysis

assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015;81:502.e1-16.
31. Chen G, Liu S, Zhao Y, *et al*. Diagnostic accuracy of endoscopic ultrasound-guided fine-needle aspiration for pancreatic cancer: A meta-analysis. *Pancreatology* 2013;13:298-304.
32. Hewitt MJ, McPhail MJ, Possamai L, *et al*. EUS-guided FNA for diagnosis of solid pancreatic neoplasms: A meta-analysis. *Gastrointest Endosc* 2012;75:319-31.
33. Facciorusso A, Bajwa HS, Menon K, *et al*. Comparison between 22G aspiration and 22G biopsy needles for EUS-guided sampling of pancreatic lesions: A meta-analysis. *Endosc Ultrasound* 2020;9:167-74.
34. Facciorusso A, Del Prete V, Buccino VR, *et al*. Diagnostic yield of Franseen and Fork-Tip biopsy needles for endoscopic ultrasound-guided tissue acquisition: A meta-analysis. *Endosc Int Open* 2019;7:E1221-30.
35. Facciorusso A, Wani S, Triantafyllou K, *et al*. Comparative accuracy of needle sizes and designs for EUS tissue sampling of solid pancreatic masses: A network meta-analysis. *Gastrointest Endosc* 2019;90:893-903.e7.
36. Li H, Li W, Zhou QY, *et al*. Fine needle biopsy is superior to fine needle aspiration in endoscopic ultrasound guided sampling of pancreatic masses: A meta-analysis of randomized controlled trials. *Medicine (Baltimore)* 2018;97:e0207.
37. Mohan BP, Shakhatreh M, Garg R, *et al*. Comparison of Franseen and Fork-Tip needles for EUS-guided fine-needle biopsy of solid mass lesions: A systematic review and meta-analysis. *Endosc Ultrasound* 2019;8:382-91.
38. Szegedy C, Ioffe S, Vanhoucke V, *et al*. Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning, In Proceedings of the AAAI Conference on Artificial Intelligence; 2017.
39. Alba AC, Alexander PE, Chang J, *et al*. High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. *J Clin Epidemiol* 2016;70:129-35.

APPENDICES

Appendix 1. Literature search strategy

Number of results before and after de-duplication		
Database	**Number of initial hits**	**After de-duplication**
EBM reviews	112	38
Embase	2260	1508
Medline	940	874
Scopus	2805	1512
Web of science	1430	313
Totals	7547	4245

EBM reviews

(((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)).ab, hw, ti.) AND ((endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab, hw, ti.) AND (("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "Chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)).ab, hw, ti.)

Embase (1974+)

(digestive system cancer/or exp esophagus cancer/or exp intestine cancer/or exp stomach cancer/or digestive system tumor/or exp esophagus tumor/or exp gastrointestinal tumor/or exp intestine tumor/or exp stomach tumor/ or ((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)).ab, kw, ti.) AND (digestive tract endoscopy/or exp chromoendoscopy/or exp endoscopic retrograde cholangiopancreatography/or exp esophagogastroduodenoscopy/or exp esophagoscopy/or exp gastrointestinal endoscopy/or digestive endoscope/or exp anoscope/or exp balloon enteroscope/or exp capsule endoscope/or exp colonoscope/or exp digestive endomicroscope/or exp duodenoscope/or exp esophagoscope/or exp gastroduodenoscope/or exp gastroscope/or exp proctoscope/or exp sigmoidoscope/or

(endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab, kw, ti.) AND (exp artificial intelligence/or exp machine learning/or ("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel

method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "Chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)).ab, kw, ti.) NOT (exp animal/not exp human/, exp child/not exp adult/, "case report".kw, pt, ti.) Limit to English

Ovid MEDLINE (R) 1946 to Present and Epub Ahead of Print, In-Process and Other Nonindexed Citations and Ovid MEDLINE (R) Daily

(exp Gastrointestinal Neoplasms/or ((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus) adj3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)).ab, kf, ti.) AND (exp Endoscopy, Digestive System/or exp Endoscopes, Gastrointestinal/or (endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*).ab, kf, ti.) AND (exp Artificial Intelligence/or ("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "Chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer adj1 heuristic*)).ab, kf, ti.) NOT (exp Animals/not Humans/, exp CHILD/not exp ADULT/, "case report".kf, pt, ti.) Limit to English

Scopus

(TITLE-ABS-KEY ((digestive OR gastr* OR gi OR alimentary OR esophag* OR oesophag* OR stomach OR intestin* OR bowel* OR colon* OR colorectal OR rectal OR rectum OR sigmoid OR duoden* OR ileum OR ileal OR jejun* OR anal OR anus) W/3 (polyp* OR mass* OR lesion* OR tumor* OR tumour* OR carcin* OR adeno* OR neoplas* OR cancer* OR malignan* OR sarcoma* OR lymphoma* OR leiomyosarcoma*))) AND (TITLE-ABS-KEY (endoscop* OR enteroscop* OR gastroscop* OR colonoscop* OR duodenoscop* OR rectoscop* OR sigmoidoscop* OR ileocolonoscop* OR chromoendoscop* OR esophagogastroduodenoscop* OR esophagoscop* OR oesophagogastroduodenoscop* OR proctoscop* OR ercp OR anoscop* OR endomicroscop* OR oesophagoscop* OR gastroduodenoscop* OR sigmoidoscop* OR diagnos* OR patholog*)) AND (TITLE-ABS-KEY ("artificial intelligence" or "machine learning" OR "machine intelligen*" OR computer-aided OR "computational intelligen*" OR "deep learning" OR "deep unified network*" OR "data mining" OR datamining OR "supervised learning" OR "semi-supervised learning" OR "unsupervised learning" OR "automated pattern recognition" OR "Bayesian learning" OR "computer heuristics" OR "hidden Markov model*" OR "k-nearest neighbor*" OR "kernel method*" OR "learning algorithm*" OR "natural language processing" OR "support vector" OR "vector machine" OR gaussian OR bootstrap OR "regression tree*" OR "linear discriminant analysis" OR "naive Bayes" OR "learning vector" OR "random forest*" OR "Chi-square automatic interaction detection" OR "iterative dichotom*" OR fuzzy OR "neural network*" OR perceptron* OR (computer AND W/1 AND heuristic*))) AND (LIMIT-TO (LANGUAGE, "English"))

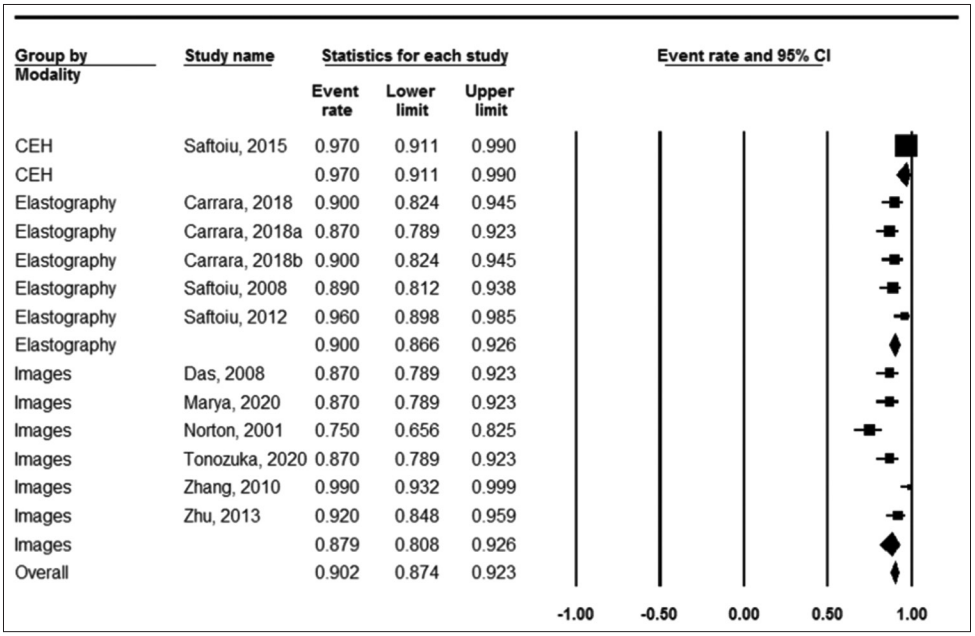
Web of Science

TS=((digestive or gastr* or GI or alimentary or esophag* or oesophag* or stomach or intestin* or bowel* or colon* or colorectal or rectal or rectum or sigmoid or duoden* or ileum or ileal or jejun* or anal or anus)

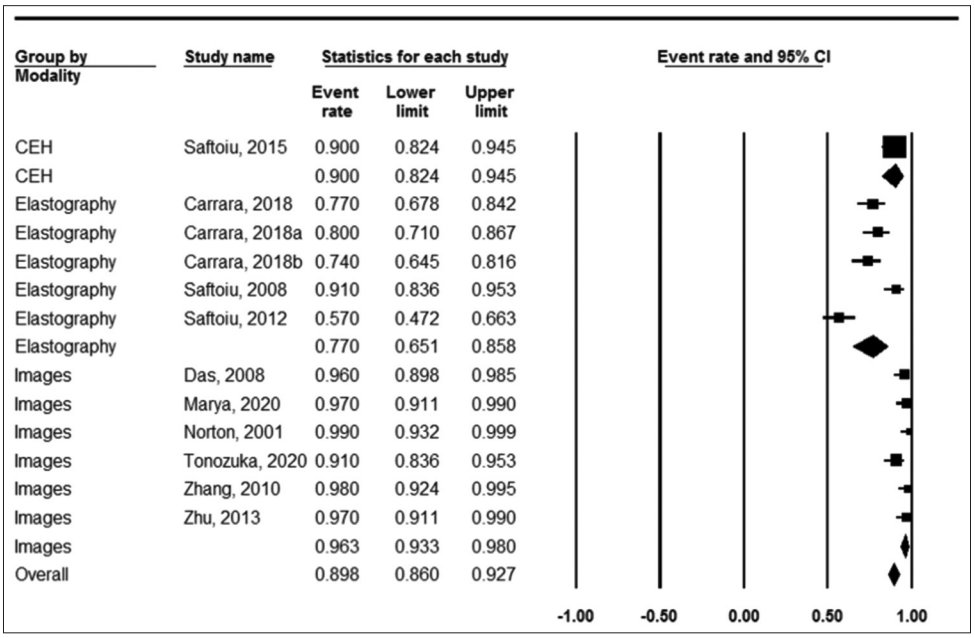
NEAR/3 (polyp* or mass* or lesion* or tumor* or tumour* or carcin* or adeno* or neoplas* or cancer* or malignan* or sarcoma* or lymphoma* or leiomyosarcoma*)) AND TS=(endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmoidoscop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or oesophagoscop* or gastroduodenoscop* or sigmoidoscop* or diagnos* or patholog*) AND TS=("artificial intelligence" or "machine learning" or "machine intelligen*" or computer-aided or "computational intelligen*" or "deep learning" or "deep unified network*" or "data mining" or datamining or "supervised learning" or "semi-supervised learning" or "unsupervised learning" or "automated pattern recognition" or "Bayesian learning" or "computer heuristics" or "hidden Markov model*" or "k-nearest neighbor*" or "kernel method*" or "learning algorithm*" or "natural language processing" or "support vector" or "vector machine" or Gaussian or Bootstrap or "regression tree*" or "linear discriminant analysis" or "naive Bayes" or "learning vector" or "random forest*" or "Chi-square automatic interaction detection" or "iterative dichotom*" or fuzzy or "neural network*" or perceptron* or (computer NEAR/1 heuristic*)) Limit to English

Appendix 2. Meta-analysis of observational studies in epidemiology checklist

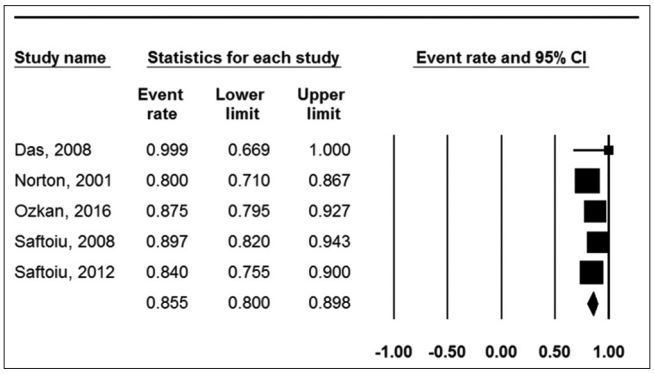
Item number	Recommendation	Reported on page number
	Reporting of background should include	
1	Problem definition	6
2	Hypothesis statement	NA
3	Description of study outcome (s)	6
4	Type of exposure or intervention used	6
5	Type of study designs used	6
6	Study population	6
	Reporting of search strategy should include	
7	Qualifications of searchers (*e.g.*, librarians and investigators)	8, Appendix 1
8	Search strategy, including time period included in the synthesis and key words	8, Appendix 1
9	Effort to include all available studies, including contact with authors	8
10	Databases and registries searched	8, Appendix 1
11	Search software used, name and version, including special features used (*e.g.*, explosion)	Appendix 1
12	Use of hand searching (*e.g.*, reference lists of obtained articles)	NA
13	List of citations located and those excluded, including justification	Appendix 1
14	Method of addressing articles published in languages other than English	8
15	Method of handling abstracts and unpublished studies	8
16	Description of any contact with authors	8
	Reporting of methods should include	
17	Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	8
18	Rationale for the selection and coding of data (*e.g.*, sound clinical principles or convenience)	8
19	Documentation of how data were classified and coded (*e.g.*, multiple raters, blinding, and inter-rater reliability)	NA
20	Assessment of confounding (*e.g.*, comparability of cases and controls in studies where appropriate)	NA
21	Assessment of study quality, including blinding of quality assessors, stratification or regression on possible predictors of study results	9
22	Assessment of heterogeneity	9
23	Description of statistical methods (*e.g.*, complete description of fixed or random-effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated	9
24	Provision of appropriate tables and graphics	Tables 1, 2, supplemental materials
	Reporting of results should include	
25	Graphic summarizing individual study estimates and the overall estimate	Figure 1, 2, 3, supplementary materials
26	Table giving descriptive information for each study included	Table 1
27	Results of sensitivity testing (*e.g.*, subgroup analysis)	11, Table 2
28	Indication of statistical uncertainty of findings	11
	Reporting of discussion should include	
29	Quantitative assessment of bias (*e.g.*, publication bias)	13
30	Justification for exclusion (*e.g.*, exclusion of non-English language citations)	NA
31	Assessment of quality of included studies	12, Supplementary Table 1
	Reporting of conclusions should include	
32	Consideration of alternative explanations for observed results	14-16
33	Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review)	14-16
34	Guidelines for future research	16



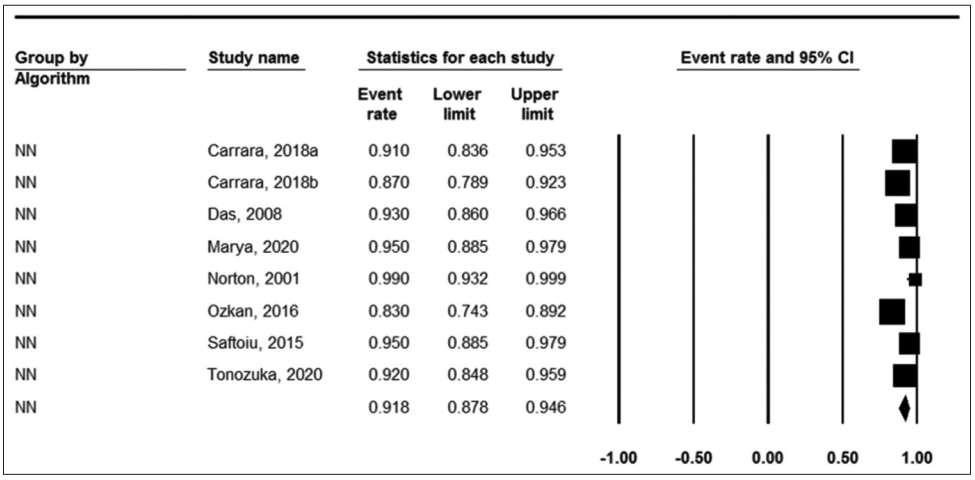
Supplementary Figure 1. Forest plot, positive predictive value. Heterogeneity: I^2% = 70%, 95% prediction interval = 65 to 97



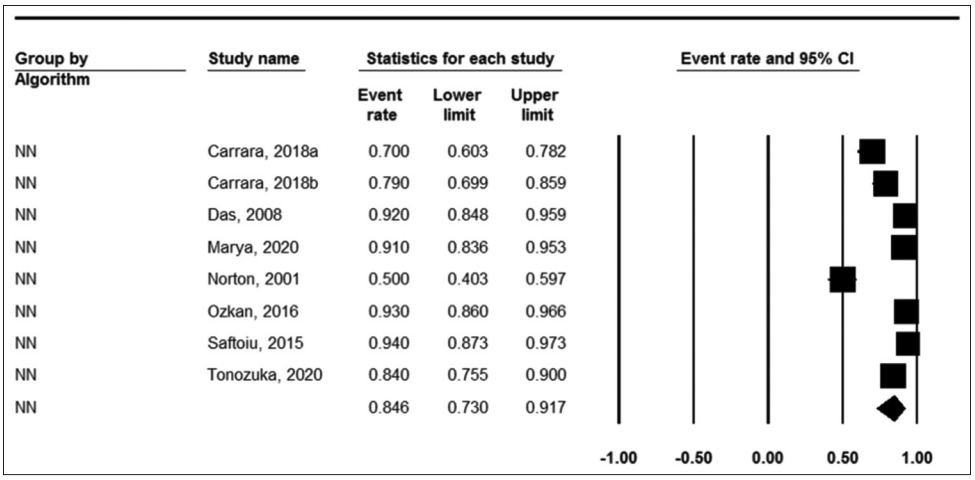
Supplementary Figure 2. Forest plot, negative predictive value. Heterogeneity: I^2% = 90%, 95% prediction interval = 51 to 99



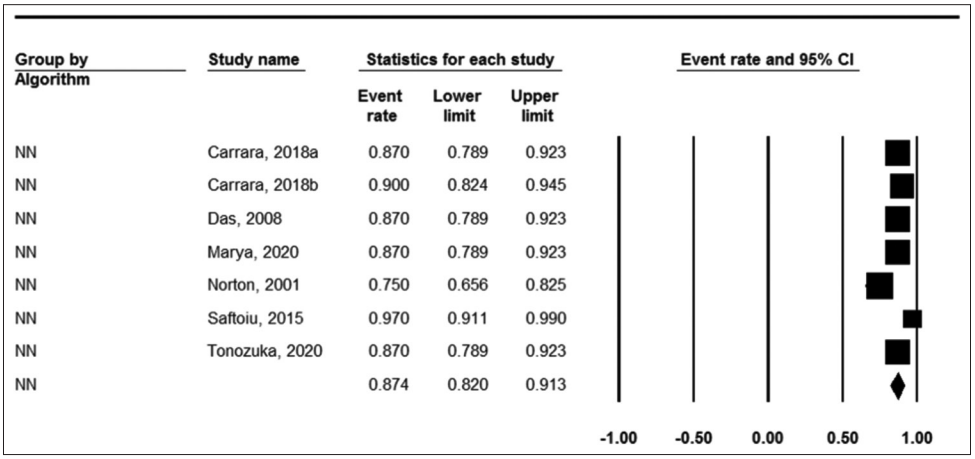
Supplementary Figure 3. Forest plot, accuracy – neural networks. Heterogeneity: I^2% = 69%, 95% prediction interval = 61 to 97



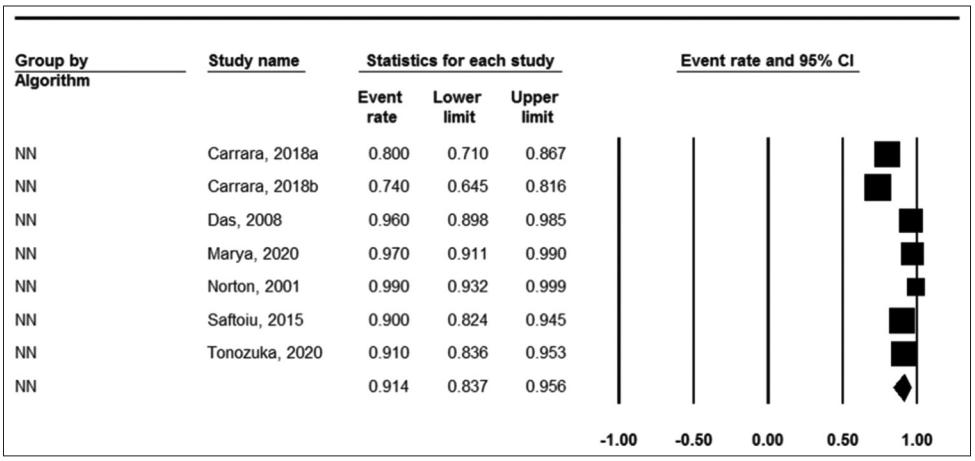
Supplementary Figure 4. Forest plot, sensitivity – neural networks. Heterogeneity: I^2% = 45%, 95% prediction interval = 84 to 97



Supplementary Figure 5. Forest plot, specificity – neural networks. Heterogeneity: I^2% = 90%, 95% prediction interval = 39 to 97



Supplementary Figure 6. Forest plot, positive predictive value – neural networks. Heterogeneity: I^2% = 68%, 95% prediction interval = 59 to 96



Supplementary Figure 7. Forest plot, negative predictive value – neural networks. Heterogeneity: I^2% = 85%, 95% prediction interval = 43 to 98

Supplementary Table 1. Quality assessment of diagnostic accuracy studies study quality assessment

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Carrara, 2018	☺	☹	☺	?	☺	☺	☺
Das, 2008	☹	☺	?	?	☺	?	?
Marya, 2020	☹	?	☹	☹	☺	☺	☺
Norton, 2001	☹	☹	☺	☹	☺	☺	☺
Ozkan, 2016	?	☺	?	?	?	☺	?
Saftoiu, 2008	?	☹	☺	?	☺	☺	☺
Saftoiu, 2012	?	☹	☺	?	☺	☺	☺
Saftoiu, 2015	☺	?	☺	☺	☺	☺	☺
Tonozuka, 2020	?	☺	☺	?	☺	☺	☺
Zhang, 2010	?	?	?	?	☺	☺	?
Zhu, 2013	?	☺	☺	?	☺	☺	☺

☺ Low risk; ☹ High risk; ? Unclear risk