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

Babu P. Mohan¹, Antonio Facciorusso², Shahab R. Khan³, Deepak Madhu⁴, Lena L. Kassab⁵, Suresh Ponnada⁶, Saurabh Chandan⁷, Stefano F. Crino⁸, Gursimran S. Kochhar⁹, Douglas G. Adler¹, Michael B. Wallace¹⁰

¹Department of Gastroenterology and Hepatology, University of Utah, Salt Lake City, Utah, USA; ²Gastroenterology Unit, University of Foggia, Foggia, Italy; ³Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA; ⁴Department of Gastroenterology, Aster MIMS, Calicut, Kerala, India; ⁵Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA; ⁶Department of Internal Medicine, Roanoke Medical Center, Roanoke, Virginia, USA; ⁷Department of Gastroenterology, CHI Creighton University Medical Center, Omaha, NE, USA; ⁸Gastroenterology and Digestive Endoscopy Unit, The Pancreas Institute, G. B. Rossi University Hospital, Verona, Italy; ⁹Department of Gastroenterology and Hepatology, Allegheny Health Network, Pittsburgh, PA, USA; ¹⁰Department of Gastroenterology and Hepatology, Sheikh Shakbout 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 I² 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

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.
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. 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). 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. 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 $F$ statistics. 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. Study by Kuwahara et al., assessed the ability of AI to predict malignancy in IPMN lesions and therefore was not included. 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, three on EUS elastography, and one on contrast-enhanced harmonic EUS. Majority of the studies evaluated the use of AI in help detecting and/or differentiating pancreatic malignancy from chronic pancreatitis. 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. 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, $F = 57%$) [Figure 2], sensitivity was 90.4% (95% CI 88.1–92.3, $F = 39%$) [Figure 3], specificity was 84% (95% CI 79.3–87.8, $F = 88%$) [Figure 4], positive predictive value was 90.2% (95% CI 87.4–92.3, $F = 70%$) [Supplementary Figure 1] and negative predictive value was 89.8% (95% CI 86–92.7, $F = 90%$) [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, $F = 69%$) [Supplementary Figure 3], sensitivity was 91.8% (95% CI 87.8–94.6, $F = 45%$) [Supplementary Figure 4], the specificity was 84.6% (95% CI 73.9–91.7, $F = 90%$), [Supplementary Figure 5] the positive predictive value was 87.4% (95% CI 82–91.3, $F = 68%$) [Supplementary Figure 6], and the negative predictive value was 91.4% (95% CI 83.7–95.6, $F = 85%$) [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.
| Study, year | Total patients | Study aim | Image type | Machine learning model | Accuracy (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|-------------|----------------|-----------|------------|------------------------|-------------------|---------------------|---------------------|-------------|-------------|
| Carrara, 2018 | 100 | Characterization of solitary pancreatic lesions from non-neoplastic tissue | EUS elastography | Neural network | 85.3 (95% CI, 78.4‑92.2) (wSR) / 84.31 (95% CI, 76.47‑90.20) (both) | 88.7 (95% CI, 83.5‑95.5) (wSR) / 86.5 (95% CI, 80.3‑92.8) (both) | 86.96 (95% CI, 83.10‑95.38) (both) | 90.60% | 90.60% |
| Das, 2008 | 56 (22 n=Normal pancreas), 12 n=Chronic pancreatitis, 22 n=Pancreatic adenocarcinoma, 1 n=Pancreatic neuroendocrine tumors | Differentiate malignancy from pancreatitis | EUS images | Neural network | 78.8 (95% CI, 74.29‑82.69) (both) | 83.5 (95% CI, 80.3‑92.8) (both) | 86.5 (95% CI, 80.3‑92.8) (both) | 90.60% | 90.60% |
| Marya, 2020 | 583 | Data on pancreatic adenocarcinoma | EUS images | Neural network | 89.70% | 91.40% | 87.70% | 94.00% | 97.90% |
| Norton, 2001 | 35 (14 n=Chronic pancreatitis), 12 n=Pancreatic adenocarcinoma | Differentiate malignancy from pancreatitis | Fractal-based quantitative analysis | Neural network | 87.50% | 83.30% | 93.30% | 93.30% | NR |
| Ozkan, 2016 | 332 (202 cancer and 130 noncancer) | Diagnosis of focal pancreatic lesions | EUS images | Neural network | 87.50% | 83.30% | 93.30% | 93.30% | NR |
| Saftoiu, 2008 | 68 (32 n=Normal pancreas), 36 n=Chronic pancreatitis, 11 n=Pancreatic adenocarcinoma, and 3 n=Pancrrectic neuroendocrine tumors | Diagnosis of focal pancreatic lesions | EUS images | Neural network | 87.70% | 91.40% | 87.70% | 94.00% | 97.90% |
| Saftoiu, 2012 | 774 | Diagnosis of focal pancreatic masses | CEH-EUS | Neural network | 89.70% | 91.40% | 87.70% | 94.00% | 97.90% |
| Tonozuka, 2020 | 920 (endosonographic images), 470 (images were independently tested) | Diagnosis of focal pancreatic cancer | EUS images | Neural network | 89.70% | 91.40% | 87.70% | 94.00% | 97.90% |
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)          | I² heterogeneity (95% PI) |
| **Accuracy**                  |                          |
| Overall                       | 86% (82.8‑88.6)          | 57% (71‑94)             |
| EUS‑images                     | 91.8% (82.3‑96.4)         | 78% (52‑99)             |
| EUS‑elastography               | 85.4% (82.8‑88.2)         | 0% (79‑89)              |
| Neural network algorithm       | 85.5% (80‑89.8)           | 69% (61‑97)             |
| **Sensitivity**                |                          |
| Overall                       | 90.4% (88.1‑92.3)         | 39% (83‑96)             |
| EUS‑images                     | 93.4% (88.9‑96.1)         | 60% (78‑98)             |
| EUS‑elastography               | 88.9% (85.8‑91.4)         | 0% (84‑93)              |
| Neural network algorithm       | 91.8% (87.8‑94.6)         | 45% (84‑97)             |
| **Specificity**                |                          |
| Overall                       | 84% (79.3‑87.8)           | 88% (51‑97)             |
| EUS‑images                     | 89.8% (76.3‑96)           | 92% (35‑99)             |
| EUS‑elastography               | 79.9% (73.5‑85.1)         | 61% (55‑93)             |
| Neural network algorithm       | 84.6% (73‑91.7)           | 90% (39‑97)             |
| **PPV**                        |                          |
| Overall                       | 90.2% (87.4‑92.3)         | 70% (65‑97)             |
| EUS‑images                     | 87.9% (80.8‑92.6)         | 75% (54‑96)             |
| EUS‑elastography               | 90% (86.6‑92.6)           | 16% (85‑95)             |
| Neural network algorithm       | 87.4% (82‑91.3)           | 68% (59‑96)             |
| **NPV**                        |                          |
| Overall                       | 89.8% (86‑92.7)           | 90% (51‑99)             |
| EUS‑images                     | 96.3% (93.3‑98)           | 37% (89‑98)             |
| EUS‑elastography               | 77% (65.1‑85.8)           | 86% (27‑96)             |
| Neural network algorithm       | 91.4% (83.7‑95.6)         | 85% (43‑98)             |
Total studies found on search of Evidence Based Medicine reviews (112), Embase (2260), Medline (940), Scopus (2805), Web of Science (1430) [Total = 7547]

After de-duplication (n = 4245)

Records screened (n = 4245)

Studies not relevant to current analysis and excluded by reading study title and abstract, n = 4206

Full-text articles assessed for eligibility (n = 39)

Studies included in quantitative synthesis (meta-analysis) (n = 11)

- Review Papers = 4
- Studies on lymph node metastases = 12
- Studies on survival prediction = 11
- Non pancreatic mass study = 1

Figure 1. Study selection flow chart

Table: Event rate and 95% CI

| Group by EUS modality | Study name | Event rate | Lower limit | Upper limit |
|----------------------|------------|------------|-------------|-------------|
| Elastography         | Carrara, 2018 | 0.853 | 0.769 | 0.910 |
| Elastography         | Carrara, 2018a | 0.843 | 0.758 | 0.902 |
| Elastography         | Carrara, 2018b | 0.843 | 0.758 | 0.902 |
| Elastography         | Saftou, 2008 | 0.897 | 0.820 | 0.943 |
| Elastography         | Saftou, 2012 | 0.840 | 0.755 | 0.900 |
| Elastography         | Das, 2008 | 0.999 | 0.669 | 1.000 |
| Images               | Norton, 2001 | 0.800 | 0.710 | 0.867 |
| Images               | Ozkan, 2016 | 0.875 | 0.795 | 0.927 |
| Images               | Zhang, 2010 | 0.980 | 0.924 | 0.995 |
| Images               | Zhu, 2013 | 0.940 | 0.873 | 0.973 |
| Images               | Overall | 0.918 | 0.823 | 0.964 |

Event rate and 95% CI

-1.00 -0.50 0.00 0.50 1.00

Heterogeneity: I² = 57%
95% prediction interval = 71 to 94

Figure 2. Forest plot, accuracy
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. 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. 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. 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. 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. 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.

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 $P$% statistics is higher when considering continuous variables as compared to categorical outcomes due to the intrinsic numeric nature of these variables. Therefore, $P$% 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 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. Magadur 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 ultrasound 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 colonooscopy versus standard colonooscopy for improving adenoma detection rate: A meta-analysis of randomized-controlled trials. EClinicalMedicine 2020;29:100622.
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;93:356-4.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;10:1-9.
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, Wiersma 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. Sältöiu A, Vilpmann P, Dietrich CF, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with video). Gastrointest Endosc 2015;82:59-69.
19. Sältöiu A, Vilpmann 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. Sältöiu A, Vilpmann 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. Mohamed 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

| 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 malignant* 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 esophagogastroduodenoscop* or esophagoscopy* or oesophagogastrroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or endoscopy* or sigmodoscop* 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 datamin 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 malignant* 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 gastroendoscopy/or exp gastrointestinal endoscopy/or exp digestive endoscopy/or exp anoscope/or exp balloon enteroscope/or exp capsule endoscopy/or exp colonoscope/or exp digestive endomicroscope/or exp duodenoscope/or exp esophagoscope/or exp gastroduodenoscope/or exp gastroscope/or exp proctoscope/or exp sigmodoscope/or

(endoscop* or enteroscop* or gastroscop* or colonoscop* or duodenoscop* or rectoscop* or sigmodoscop* or ileocolonoscop* or chromoendoscopy* or esophagogastroduodenoscop* or esophagoscope* or oesophagogastrroduodenoscop* or proctoscop* or ERCP or anoscop* or endomicroscop* or esophagoscope* or gastroduodenoscop* or sigmodoscop* 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 datamin 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 sigmoidsocop* or ileocolonoscop* or chromoendoscop* or esophagogastroduodenoscop* or esophagoscop* or oesophagogastrodudosenoscop* 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 intelligent” or “deep learning” or “deep unified network*” or “data mining” or datamin* 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 exp 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 carcino* OR sarcoma* OR lymphoma* OR leiomyosarcoma*))) AND (TITLE-ABS-KEY (endoscop* OR enteroscop* OR gastroscop* OR colonoscop* OR duodenoscop* OR rectoscop* OR sigmoidsocop* OR ileocolonoskop* OR chromoendoscop* OR esophagogastroduodenoscop* OR esophagoscop* OR oesophagogastrodudosenoscop* OR proctoscop* OR ercp OR anoscop* OR endomicroscop* OR oesophagoskop* OR gastroduodenoscop* OR sigmoidoscop* OR diagnos* OR patholog*)) AND (TITLE-ABS-KEY (“artificial intelligence” or “machine learning” or “machine intelligent*” or computer-aided or “computational intelligent*” or “deep learning” or “deep unified network*” or “data mining” or datamin* 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 gastroskop* or colonoskop* or duodenoskop* or rectoskop* or sigmoidoskop* or ileocolonoskop* or chromoendoskop* or esophagogastroduodenoskop* or esophagoskop* or oesophagogastroduodenoskop* or proctoskop* or ERCP or anoskop* or endomicroskop* or oesophagoskop* or gastroduodenoskop* or sigmoidoskop* or diagnos* or patholog*) AND TS=(“artificial intelligence” or “machine learning” or “machine intelligence” or computer-aided or “computational intelligence” 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 |
|-------------|---------------------------------------------------------------------------------|-------------------------|
| 1           | Reporting of background should include                                           |                         |
| 2           | Problem definition                                                              | 6                       |
| 3           | Hypothesis statement                                                            | NA                      |
| 4           | Description of study outcome(s)                                                 | 6                       |
| 5           | Type of exposure or intervention used                                           | 6                       |
| 6           | Type of study designs used                                                       | 6                       |
| 7           | Reporting of search strategy should include                                     |                         |
| 8           | Study population                                                                | 6                       |
| 9           | Reporting of search strategy should include                                     |                         |
| 10          | Qualifications of searchers (e.g., librarians and investigators)                | Appendix 1              |
| 11          | Search strategy, including time period included in the synthesis and key words  |                         |
| 12          | Effort to include all available studies, including contact with authors          | 8                       |
| 13          | Databases and registries searched                                               | APPENDIX 1              |
| 14          | Search software used, name and version, including special features used (e.g., explosion) |                         |
| 15          | Use of hand searching (e.g., reference lists of obtained articles)              | NA                      |
| 16          | List of citations located and those excluded, including justification           | APPENDIX 1              |
| 17          | Method of addressing articles published in languages other than English         | 8                       |
| 18          | Method of handling abstracts and unpublished studies                            | 8                       |
| 19          | Description of any contact with authors                                         | 8                       |
| 20          | Reporting of methods should include                                             |                         |
| 21          | Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested | 8                       |
| 22          | Rationale for the selection and coding of data (e.g., sound clinical principles or convenience) | 8                       |
| 23          | Documentation of how data were classified and coded (e.g., multiple raters, blinding, and inter-rater reliability) | NA                      |
| 24          | Assessment of confounding (e.g., comparability of cases and controls in studies where appropriate) | NA                      |
| 25          | Assessment of quality, including blinding of quality assessors, stratification or regression on possible predictors of study results | 9                       |
| 26          | Assessment of heterogeneity                                                     | 9                       |
| 27          | 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                       |
| 28          | Provision of appropriate tables and graphics                                     | Tables 1, 2, supplemental materials |
| 29          | Reporting of results should include                                             |                         |
| 30          | Graphic summarizing individual study estimates and the overall estimate         | Figure 1, 2, 3, supplementary materials |
| 31          | Table giving descriptive information for each study included                    | Table 1                 |
| 32          | Results of sensitivity testing (e.g., subgroup analysis)                         | 11, Table 2             |
| 33          | Indication of statistical uncertainty of findings                               | 11                      |
| 34          | Reporting of discussion should include                                          |                         |
| 35          | Quantitative assessment of bias (e.g., publication bias)                        | 13                      |
| 36          | Justification for exclusion (e.g., exclusion of non-English language citations) | NA                      |
| 37          | Assessment of quality of included studies                                       | 12, Supplementary Table 1|
| 38          | Reporting of conclusions should include                                         |                         |
| 39          | Consideration of alternative explanations for observed results                  | 14-16                   |
| 40          | Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review) | 14-16                   |
| 41          | 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

| Study name | Statistics for each study | Event rate and 95% CI |
|------------|---------------------------|-----------------------|
|            | Event rate | Lower limit | Upper limit |
| Das, 2008  | 0.999      | 0.669       | 1.000       |
| Norton, 2001 | 0.800     | 0.710       | 0.867       |
| Ozkan, 2016 | 0.875      | 0.795       | 0.927       |
| Saftoiu, 2008 | 0.897     | 0.820       | 0.943       |
| Saftoiu, 2012 | 0.840     | 0.755       | 0.900       |
|            | 0.855      | 0.800       | 0.898       |

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

| Group by Algorithm | Study name       | Statistics for each study | Event rate and 95% CI |
|-------------------|------------------|---------------------------|-----------------------|
|                   | Event rate | Lower limit | Upper limit |
| NN                | Carrara, 2018a  | 0.910 | 0.836 | 0.953 |
| NN                | Carrara, 2018b  | 0.870 | 0.789 | 0.923 |
| NN                | Das, 2008       | 0.930 | 0.860 | 0.966 |
| NN                | Marya, 2020     | 0.950 | 0.885 | 0.979 |
| NN                | Norton, 2001    | 0.990 | 0.932 | 0.999 |
| NN                | Ozkan, 2016     | 0.830 | 0.743 | 0.892 |
| NN                | Saftoiu, 2015   | 0.950 | 0.885 | 0.979 |
| NN                | Tonozuka, 2020  | 0.920 | 0.848 | 0.959 |
| NN                |                 | 0.918 | 0.878 | 0.946 |

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

| Group by Algorithm | Study name       | Statistics for each study | Event rate and 95% CI |
|-------------------|------------------|---------------------------|-----------------------|
|                   | Event rate | Lower limit | Upper limit |
| NN                | Carrara, 2018a  | 0.700 | 0.603 | 0.782 |
| NN                | Carrara, 2018b  | 0.790 | 0.699 | 0.859 |
| NN                | Das, 2008       | 0.920 | 0.848 | 0.959 |
| NN                | Marya, 2020     | 0.910 | 0.836 | 0.953 |
| NN                | Norton, 2001    | 0.500 | 0.403 | 0.597 |
| NN                | Ozkan, 2016     | 0.930 | 0.860 | 0.966 |
| NN                | Saftoiu, 2015   | 0.940 | 0.873 | 0.973 |
| NN                | Tonozuka, 2020  | 0.840 | 0.755 | 0.900 |
| NN                |                 | 0.846 | 0.730 | 0.917 |

-1.00  -0.50  0.00  0.50  1.00
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](image)

![Proportion of studies with low, high, or unclear RISK of BIAS](image)

![Proportion of studies with low, high, or unclear CONCERNS regarding APPLICABILITY](image)