Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: a systematic review and meta-analysis

Lei Liu,[a,g] Jiangang Liu,[b,g] Qing Su,[c] Yuening Chu,[d] Hexia Xia,[a,*] and Ran Xu[e,f,**]

[a]Department of Gynecology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, 200011, China
[b]Department of Obstetrics and Gynecology, Puren Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, 430080, China
[c]Department of Obstetrics and Gynecology, The Fourth Hospital of Changsha, Changsha, 410006, China
[d]Department of Obstetrics and Gynecology, Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai, 201204, China
[e]Department of Obstetrics and Gynecology, Affiliated Zhejiang Hospital, Zhejiang University School of Medicine, Hangzhou, 310013, China
[f]Heidelberg University, Heidelberg, 69120, Germany

eClinicalMedicine
2025;80: 102992
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102992

Summary

Background Cervical cytology screening and colposcopy play crucial roles in cervical intraepithelial neoplasia (CIN) and cervical cancer prevention. Previous studies have provided evidence that artificial intelligence (AI) has remarkable diagnostic accuracy in these procedures. With this systematic review and meta-analysis, we aimed to examine the pooled accuracy, sensitivity, and specificity of AI-assisted cervical cytology screening and colposcopy for cervical intraepithelial neoplasia and cervical cancer screening.

Methods In this systematic review and meta-analysis, we searched the PubMed, Embase, and Cochrane Library databases for studies published between January 1, 1986 and August 31, 2024. Studies investigating the sensitivity and specificity of AI-assisted cervical cytology screening and colposcopy for histologically verified cervical intraepithelial neoplasia and cervical cancer and a minimum of five cases were included. The performance of AI and experienced colposcopists was assessed via the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) through random effect models. Additionally, subgroup analyses of multiple diagnostic performance metrics in developed and developing countries were conducted. This study was registered with PROSPERO (CRD42024534049).

Findings Seventy-seven studies met the eligibility criteria for inclusion in this study. The pooled diagnostic parameters of AI-assisted cervical cytology via Papanicolaou (Pap) smears were as follows: accuracy, 94% (95% CI 92–96); sensitivity, 95% (95% CI 91–98); specificity, 94% (95% CI 89–97); PPV, 88% (95% CI 78–96); and NPV, 95% (95% CI 89–99). The pooled accuracy, sensitivity, specificity, PPV, and NPV of AI-assisted cervical cytology via ThinPrep cytologic test (TCT) were 90% (95% CI 85–94), 97% (95% CI 95–99), 94% (95% CI 85–98), 84% (95% CI 64–98), and 96% (95% CI 94–98), respectively. Subgroup analysis revealed that, for AI-assisted cervical cytology diagnosis, certain performance indicators were superior in developed countries compared to developing countries. Compared with experienced colposcopists, AI demonstrated superior accuracy in colposcopic examinations (odds ratio (OR) 1.75; 95% CI 1.33–2.31; $P < 0.0001$; $I^2 = 93\%$).

Interpretation These results underscore the potential and practical value of AI in preventing and enabling early diagnosis of cervical cancer. Further research should support the development of AI for cervical cancer screening, including in low- and middle-income countries with limited resources.

Funding This study was supported by the National Natural Science Foundation of China (No. 81901493) and the Shanghai Pujiang Program (No. 21PJD006).

*Corresponding author.
**Corresponding author. Department of Obstetrics and Gynecology, Affiliated Zhejiang Hospital, Zhejiang University School of Medicine, Hangzhou, 310013, China.
E-mail addresses: hexia_xia@fudan.edu.cn (H. Xia), ran.xu@medma.uni-heidelberg.de (R. Xu).
[g]These authors contribute equally to this work.

Keywords: Performance; Artificial intelligence; Cervical intraepithelial neoplasia; Cervical cancer; Meta-analysis

Research in context

Evidence before this study

We searched the PubMed, Embase, and Cochrane Library databases to identify studies published in English between January 1, 1986 and August 31, 2024, using the terms "cervical cancer" OR "cervical intraepithelial neoplasia" AND "artificial intelligence". Only one review has evaluated the performance of AI systems in the prediction, screening, or detection of cervical cancer and precancerous lesions. However, to our knowledge, no previous systematic review and meta-analysis has estimated the pooled diagnostic accuracy (sensitivity and specificity) of AI-assisted cervical cytology screening; similarly, no systematic review and meta-analysis have assessed the performance of AI compared with colposcopists for the detection of CIN or cervical cancer via colposcopy.

Added value of this study

This meta-analysis indicated that AI has acceptable performance in distinguishing between normal and abnormal cytological results in cervical cytology screening in both developing and developed countries. Similarly, AI exhibited higher accuracy in colposcopic examinations compared with experienced colposcopists in both settings. Furthermore, AI outperformed experienced colposcopists in diagnosing LSIL+ and HSIL+ patients.

Implications of all the available evidence

These results highlight the potential and application value of AI in the prevention and early diagnosis of cervical cancer, also in LMICs where there is a lack of experienced cytopathologists and colposcopists.

Introduction

Cervical cancer is a prominent contributor to both the morbidity and mortality rates associated with cancer in women globally. Annually, approximately 600,000 women are diagnosed with cervical cancer worldwide, with over 300,000 succumbing to the disease.[1] Over 85% of new cervical cancer cases and 87% of deaths due to cervical cancer take place in low-income and middle-income countries (LMICs). Women residing in LMICs, which have the highest cervical cancer incidence and mortality rates, bear a disproportionately high burden of cervical cancer.[1,2] On November 17, 2020, the World Health Organization (WHO) initiated a global endeavor aimed at eradicating cervical cancer as a significant public health issue. The WHO has proposed a worldwide threshold of elimination, which is set at four cases per 100,000 person-years, and advocates for the implementation of a comprehensive triple intervention strategy. This strategy encompasses the vaccination of a minimum of 90% of girls against human papillomavirus (HPV) before they reach 15 years of age, screening 70% of women utilizing a high-performance test by 35 years of age and subsequently at 45 years of age, and ensuring treatment for at least 90% of detected precancerous lesions and invasive cancers.[3–5] Cervical cancer is preventable through vaccination and can be effectively treated when it is diagnosed at an early stage. The recent surge in the adoption of HPV vaccination has predominantly been observed in high-income nations. In contrast to high-income countries, in which the coverage rate exceeds 85%, fewer than 30% of LMICs have implemented HPV vaccination initiatives.[6] Efficient screening for precancerous cervical lesions is the sole protective intervention available to women who have not received vaccination.[7] Currently, screening methods primarily include HPV testing, cervical cytology, and DNA ploidy analysis. Because of its simplicity and cost-effectiveness, cervical cytology screening is recommended for population-based screening programs. For individuals with abnormal findings on screening tests, the diagnostic procedure typically entails the utilization of colposcopy in conjunction with biopsy to identify precancerous lesions. Subsequent therapeutic interventions are employed to prevent the development of cancer. The accurate identification of aberrant cervical epithelium and systematic performance of targeted biopsies on all acetowhite regions during colposcopic examination are critical to prevent overlooking of high-grade squamous intraepithelial lesions or worse (HSIL+) that require immediate treatment.[8,9] Manual review of cervical cytology slides, whether from traditional smears or liquid-based preparations, is labor-intensive, prone to human error, and highly dependent on the expertise of the cytopathologist, potentially leading to reduced sensitivity and an increased likelihood of false negatives.[10] In addition, the existing colposcopic evaluation methods pose notable challenges, especially in LMICs. These challenges encompass inadequate concordance (less than 50%) between colposcopy result interpretations and pathological outcomes, a pronounced reliance on the subjective experience of operators, considerable disparities among inter- and intra-operators, high rates of unnecessary colposcopies, a

high percentage of multiple and random punch biopsies, a scarcity of proficient colposcopists, and deficiencies in the implementation of quality control and quality assurance measures.[11–15]

Artificial intelligence (AI) constitutes a subfield of computer science in which algorithms perform tasks that are conventionally executed by humans. Machine learning (ML) is a term that refers to a group of techniques in the field of AI that allow algorithms to learn from data, iteratively improving their own performance without the need for explicit programming.[16] The number of AI algorithms within the medical literature and healthcare industry is increasing rapidly. This increase is attributed to recent advancements in computational power, data accessibility, and model complexity through research in mathematics and computer science.[17] In alignment with this trend, there has been a notable increase in the approval of AI algorithms and medical devices with AI capabilities by the U.S. Food and Drug Administration.[18] The process of medical AI research is often the same: starting with an image classification task, training an AI system via supervised learning on labeled data, and then assessing the system by comparing it to human specialists.[19] AI has been used to perform a variety of medical image analysis tasks with performance on par with that of clinical experts. These tasks include the identification of pulmonary diseases, the identification and grading of diabetic retinopathy, the classification of skin lesions, the differentiation of benign lesions from malignant lesions, such as lung cancer,[20] breast cancer,[21] pancreatic cancer,[22] and renal cancer,[23] and so forth.[24,25] Previous studies have shown exceptional accuracy of AI in pathological differentiation and have focused on AI-enabled diagnostic approaches for discriminating between low-grade squamous intraepithelial lesions (LSIL) and HSIL.[26–31] In addition, considerable advancements have been achieved in the development of AI-based digital colposcopy aimed at enhancing the effectiveness and precision of clinical diagnoses, and multiple preliminary studies of computer algorithms applied to cervical images have been carried out.[31–35] However, some studies have shown that AI can match or surpass colposcopists in terms of the sensitivity and specificity of colposcopy in detecting cervical intraepithelial neoplasia (CIN) and cervical cancer during primary cervical cancer screening, whereas others have reported opposite findings.[9,29,36,37] Published reviews have assessed the performance of AI compared with that of healthcare professionals in detecting cervical precancer or cancer via medical imaging.[38–41] However, a meta-analysis of test accuracy has not yet been performed.

The objective of this systematic review and meta-analysis was to assess existing evidence of the diagnostic accuracy, sensitivity, and specificity of AI-assisted cervical cytology screening and colposcopy for detecting cervical intraepithelial neoplasia and cervical cancer. The findings of this review are essential for policymakers and stakeholders involved in formulating guidelines and recommendations for cervical cancer screening in LMICs, thereby advancing global efforts towards the eventual elimination of cervical cancer.

Methods

Search strategy and selection criteria

A systematic review of the literature on the concepts of AI-assisted cervical cytology screening and colposcopy for the detection of cervical intraepithelial neoplasia and cervical cancer was conducted. The MEDLINE, Embase, and Cochrane Library databases were searched for all articles published from database inception to August 31, 2024. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 reporting guidelines were used to design the review. The inclusion criteria and analysis plan were decided a priori and registered on PROSPERO (CRD42024534049). A manual reference search of the included articles was also performed to identify any additional relevant articles. We included studies that developed and applied machine learning (ML) or deep learning models for classifying cervical precancer lesions using medical imaging.

Data extraction

We included all types of observational studies associated with AI-assisted cervical cytology screening and colposcopy for the detection of CIN and cervical cancer that were performed anywhere in the world and published exclusively in English. Studies focused on animals, gene expression profiling, genomic and molecular studies, chromosomal alteration progression, nuclei segmentation, genomic profiles, spectroscopy, optoelectronic sensors, biomarkers, cervical cancer prognosis, and mathematical models were excluded. In addition, case reports, literature reviews, abstracts, letters to editors, non-full-text articles, and unpublished studies were excluded. The screening of the studies on the basis of titles and abstracts was performed by two independent reviewers (LL and QS), with conflicts resolved by consulting a third reviewer (RX). Three reviewers independently assessed eligibility for inclusion following abstract screening according to the inclusion and exclusion criteria. In cases of conflict, decisions were made through consensus agreement among the three reviewers. Eligible full-text articles were evaluated, and relevant data were extracted independently by two reviewers (LL and QS) via a template data extraction form, with conflicts resolved by the inclusion of a third reviewer (RX).

For studies that generated multiple AI models to classify the same outcome, we included the performance data from the model that performed best on the

test set. This ensured that only the most effective model would be considered for clinical application. When data were available, we compared the accuracy of colposcopists in classifying CIN and cervical cancer to that of the AI models, excluding the performance of junior colposcopists. If multiple colposcopists were evaluated, we calculated their mean accuracy score on the basis of their performance in classifying CIN and cervical cancer. A mean score was calculated for colposcopist performance because this score most closely resembled the diagnostic accuracy of the current workflow.

The following data were extracted from eligible studies: first author, year of publication, country in which data were collected, study type, number of images used for training, validation and testing, AI algorithm type, database used, data availability, and number of patients who underwent both a trial of AI screening and a clinician reference test. Diagnostic performance data were extracted to estimate the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) if data on the number of patients who had true positive (TP), false positive (FP), false negative (FN), and true negative (TN) results were provided. For studies lacking complete information for the meta-analysis, we contacted the authors to obtain the necessary data. When data were unavailable, we incorporated as much information as possible from the study into the narrative synthesis.

Assessment of study bias

We evaluated the risk of bias for each study via the quality assessment of diagnostic accuracy studies (QUADAS-2) tool (University of Bristol, Bristol, UK), which is the recommended instrument for assessing primary studies of diagnostic accuracy in systematic reviews.[42] The risk of bias and concerns about applicability were assessed across four domains: patient selection, the index test, the reference standard, and patient flow and timing. The degree of risk or concern was categorized as high, low, or unclear.

Statistics

We extracted diagnostic performance metrics, including the reported accuracy, sensitivity, specificity, PPV, and NPV. When original data cannot be extracted from a study, we will consider its stratified data as independent research, given the study's objective to provide an overview of pooled rates from various studies rather than precise point estimates. A formal quality assessment of the studies was not conducted because of their nonclinical nature. The meta-analysis was conducted via RevMan (version 5.3) and Stata (version 18.0) to calculate pooled estimates for each case via a random effects model.[43] Heterogeneity among study-specific estimates was assessed via Cochran's Q test for heterogeneity, the 95% prediction interval to address the dispersion of effects, and the I^2 statistic.[44,45] We plotted a summary receiver operating characteristic curve to estimate the diagnostic accuracy of AI trials by examining the area under the curve and the summary operating point. To detect publication bias in diagnostic performance studies, we used Deeks' funnel plot. Subgroup and sensitivity analyses were subsequently performed. For the subgroup analysis, stratified methods and univariate meta-regression were employed, stratifying data by developed versus developing countries and different diagnostic cut-off points. Sensitivity analyses were conducted via the leave-one-out method. For cytology, the subgroups included Pap smears and TCT, as well as studies from developed and developing countries. For colposcopy, the subgroups consisted of AI-assisted colposcopy and clinician-performed colposcopy, categorized by developed and developing countries, and further differentiated by LSIL+ and HSIL+. The regression analysis encompassed the assessment of heterogeneity related to AI in studies on TCT, Pap smears, and colposcopy.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We identified 1488 articles from the electronic searches, which included cytology and colposcopy studies. After removing duplicates, title and abstract screening was performed for 1350 studies, with 138 studies remaining after screening. After a full-text review, 77 articles (51 cytology articles and 26 colposcopy articles) were included in the systematic review. A PRISMA flowchart of the included studies is displayed in Fig. 1. The 51 eligible cytology studies[30,46–60,61–95] were published between 2012 and 2024 and included over 280,000 tests from 17 countries in Europe, South America, North America, and Asia (Tables 1 and 2). The 26 eligible colposcopy studies[9,28,29,32,33,36,37,96–114] were published between 2012 and 2024 and included over 45,000 tests from nine countries in Europe, South America, North America, and Asia (Supplementary Table S1).

In research on AI applications in cytology, both Pap smears and ThinPrep cytologic test (TCT) were included. Some studies focusing on Pap smears utilized the traditional Pap classification system, whereas others employed the Bethesda system (TBS). Despite these variations, the results were categorized as abnormal or normal on the basis of the three-step cervical lesion screening method. For TCT studies, TBS was universally applied. Some AI studies on colposcopy adopted the CIN classification system, whereas others used TBS for diagnosing cervical lesions. Ultimately, cervical lesions were described via TBS, as LSIL corresponds to CIN1, and HSIL corresponds to CIN2 and CIN3. Although different classification systems were used,

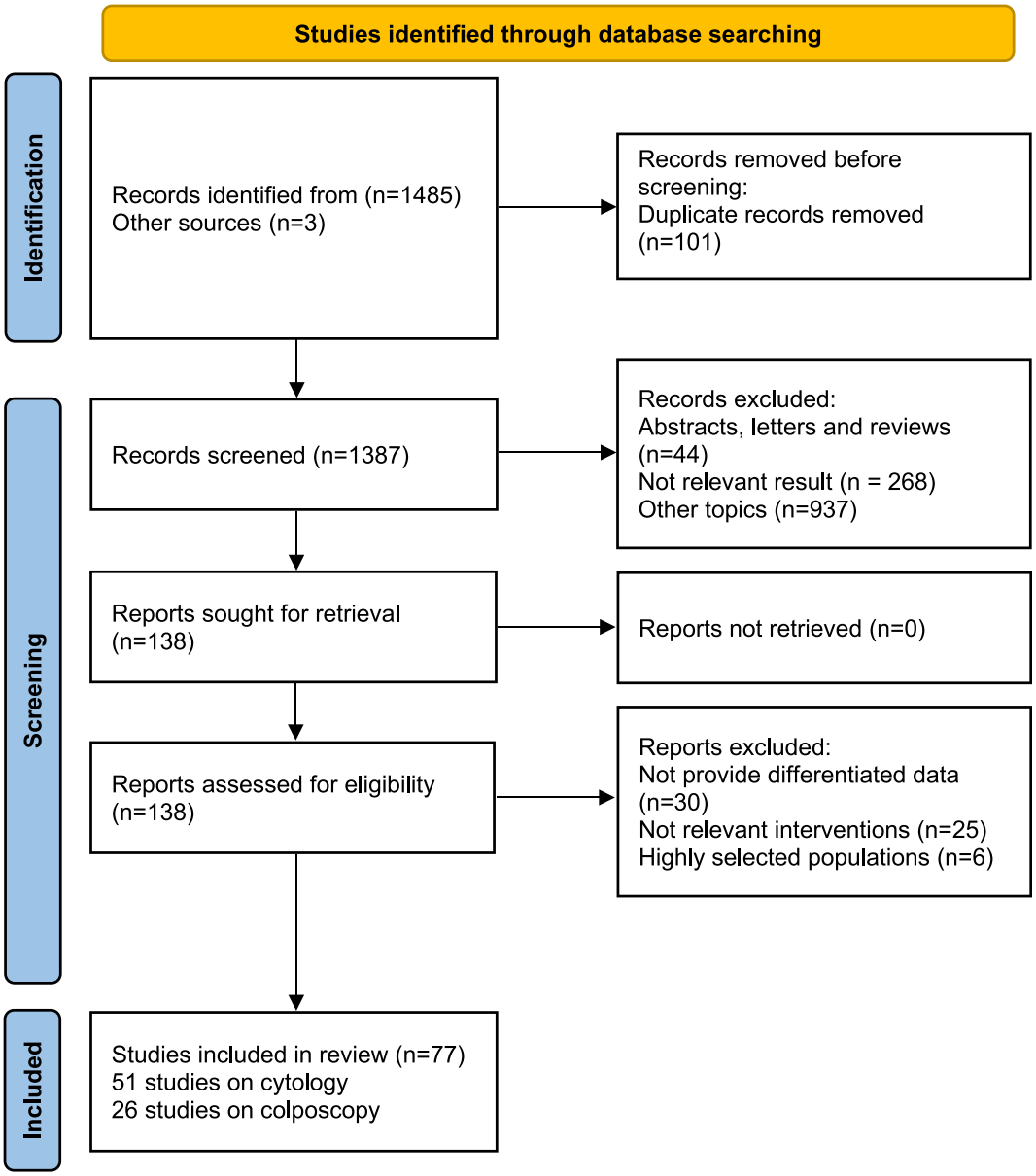

Fig. 1: The PRISMA flow diagram of literature screening. The period from January 1, 1986 to August 31, 2024, is covered. Given the pivotal advancements in artificial intelligence in 2012, particularly with the breakthroughs in deep learning technology, we have chosen to exclude studies conducted prior to this year. This decision is based on the substantial progress AI has made since 2012, especially in processing large-scale data and complex pattern recognition, where the introduction and refinement of deep learning have been transformative. Thus, to ensure that our review and analysis accurately reflect the current capabilities and performance of AI, we have included only studies published from 2012 onward.

the primary goal of this study was to evaluate AI performance, making the presence of heterogeneity acceptable. In terms of the diagnostic performance of AI-assisted cytology, data for Pap smear accuracy were extracted from 20 studies, and data for the sensitivity, specificity, PPV, and NPV were extracted from 11 studies. For TCTs, data for diagnostic accuracy were extracted from 28 studies, and data for the sensitivity, specificity, PPV, and NPV were extracted from 25 studies. For colposcopy, data for AI accuracy were collected from 26 studies, and data for the sensitivity, specificity, PPV, and NPV were extracted from 16 studies. For human experts in colposcopy, data for accuracy were extracted from 14 studies, and data for the sensitivity, specificity, PPV, and NPV were extracted from 8 studies.

Study	Country	Algorithm	Dataset	Study design	No. of images	Training size	Validation size	Testing size	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC
Al-Batah et al. (2014)[46]	Jordan	ANFIS (ML)	Royal Rehabilitation Center in King Hussein Medical Center	Retrospective	500	NR	NR	NR	NR	NR	NR	NR	NR
Hyeon et al. (2017)[90]	Korea	SVM (ML)	Seegene Medical Foundation	Retrospective	16,746	80%	–	20%	78.0	78.0	78.2	78.2	NR
Arya et al. (2018)[82]	India	ANN (ML)	DTU/Herlev Pap smear benchmark data set + Rajasthan University, Jaipur (MNITJ)	Retrospective	330	NR	NR	NR	99.0	99.0	98.8	98.8	NR
Sompawong et al. (2019)[47]	Thailand	Mask R-CNN (CNN)	Thammasat University (TU) Hospital.	Retrospective	1024	NR	NR	NR	72.5	94.3	NR	NR	NR
Hussain et al. (2020)[48]	India	DL	Babina Diagnostic Pvt. Ltd, Imphal; Gauhati Medical College and Hospital	Retrospective	1670	NR	NR	NR	97.8	97.9	NR	NR	NR
Sanyal et al. (2020)[49]	India	CNN	East India hospital	Retrospective	1838	1397	441	441	94.3	96.0	91.7	68.3	NR
Win et al. (2020)[83]	Greece	RF, LD, SVM, KNN, boosted trees, and bagged trees (ML)	SIPaKMeD	Retrospective	966	NR	NR	NR	NR	NR	NR	NR	NR
Holmstrom et al. (2021)[50]	Kenya	DLS (CNN)	Smears of HIV-positive women	Retrospective	740	350	390	361	95.7	84.7	48.4	99.3	NR
Ali et al. (2021)[51]	Multi-country	RT (ML)	Kaggle and UCI ML Repository	Retrospective	NR	NR	NR	NR	99.1	90.4	NR	NR	NR
Diniz et al. (2021)[86]	China	DT + NV + KNN (ML)	ISBI'14 Overlapping Cervical Cytology Image Segmentation Challenge dataset	Retrospective	–	45	–	900	99.9	NR	NR	NR	NR
Lin et al. (2021)[52]	China	RRS-0.85 (CNN)	Four medical centers	Retrospective	19,303	13,486	2486	3331	90.7	80.0	42.2	98.2	0.925
Sheela et al. (2021)[85]	Denmark	Deep auto encoder-based ELM (CNN)	Herlev	Retrospective	917	NR	NR	699	99.8	98.0	98.9	99.6	NR
Bhatt et al. (2021)[53]	Greece	EfficientNet-B3 (CNN)	SIPaKMeD	Retrospective	4049	NR	NR	NR	98.9	NR	NR	NR	NR
Gao et al. (2022)[54]	Greece	3cDe-Net (CNN)	SIPaKMeD cervical cell dataset	Retrospective	966	3	1	1	98.5	98.7	NR	NR	NR
Kupas et al. (2022)[55]	Hungary	DenseNet (CNN)	A private dataset	Retrospective	3005	2404	NR	601	95.5	91.0	90.8	95.6	NR
Wang et al. (2022)[56]	Denmark	3cDe-Net (CNN)	Herlev	Retrospective	917	8	1	1	NR	NR	NR	NR	NR
Alsubai et al. (2023)[57]	Greece	SIPaKMeD (CNN)	SIPaKMeD	Retrospective	4049	2832	608	609	91.5	90.9	87.2	94.0	NR
Chowdary et al. (2023)[58]	Greece	SE-UNet (CNN)	SIPAKMED	Retrospective	4049	6	2	2	99.2	99.8	99.4	99.4	NR
Kalbhor et al. (2023)[59]	India	Resnet-50 (CNN)	Sipakmed dataset	Retrospective	4049	NR	NR	NR	95	96	94.2	96.6	NR
Tan et al. (2024)[60]	Denmark	DenseNet-201 (CNN)	Herlev	Retrospective	917	8	2	1	85.4	91.5	96.7	69.3	NR
Yang et al. (2024)[61]	China	FPNC (CNN)	CCID (private)	Retrospective	148,762	7	1	2	99.5	NR	99.5	NR	NR

NR, not report; Se, sensitivity; Sp, specificity.

Table 1: **Characteristics of included studies about Pap smears.**

The pooled diagnostic parameters of AI-assisted Pap smears were as follows: accuracy, 94% (95% CI 92–96; I^2 = 99.7%); sensitivity, 95% (95% CI 91–98; I^2 = 99.8%); specificity, 94% (95% CI 89–97; I^2 = 99.8%); PPV, 88% (95% CI 78–96; I^2 = 99.4%); and NPV, 95% (95% CI 89–99; I^2 = 99.4%) (Figs. 2 and 3; Supplementary Table S2). The pooled accuracy of AI-assisted cervical cytology via TCT was 90% (95% CI 85–94; I^2 = 100%),

Study	Country	Algorithm	Dataset	Study design	No. of images	Training size	Validation size	Testing size	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC
Kim et al. (2015)[89]	Korea	HTEA Hough transform extraction algorithm (ML)	NR	Retrospective	NR	NR	NR	NR	NR	NR	NR	NR	NR
Kyrgiou et al. (2016)[62]	UK	MLP (DL)	Three University Hospitals	Prospective + Retrospective	3651	NR	NR	NR	99.47	99.51	98.95	99.75	NR
Lasyk et al. (2019)[63]	Porland	U-NET and CNN	Pomeranian Medical University in Szczecin	Retrospective	2058	NR	NR	NR	100	100	100	100	NR
Sanyal et al. (2019)[64]	India	CNN	A tertiary care hospital	Retrospective	2816	820	400	1596	95.63	79.85	41.3	99.2	NR
Xiang et al. (2020)[88]	China	YOLOv3 (CNN)	Central South University (own dataset)	Retrospective	1014	NR	NR	NR	97.5	67.8	88.53	91.51	89.3
Tan et al. (2021)[30]	China	Robust DCNN model (CNN)	Multiple collaborating hospitals	Retrospective	16,366	13,775	2301	290	99.4	34.8	70.8	97.5	NR
Tang et al. (2021)[65]	China	RetinaNet (CNN)	Shenzhen Maternity and Child Healthcare Hospital	Retrospective	2167	NR	NR	1944	94.56	89.55	92.77	92.07	NR
Cao et al. (2021)[66]	China	AttFPN (CNN)	HMUCH dataset + HMCHH dataset	Retrospective	NR	NR	NR	3970	95.83	94.81	98.2	88.51	99.1
Jia et al. (2021)[86]	NR	SSD (CNN)	NR	Retrospective	1462	1167	NR	295	95.7	89.9	NR	NR	NR
Li et al. (2021)[67]	China	RCNN-FPN (CNN)	Alibaba Cloud TianChi Company	Retrospective	800	640	NR	160	NR	NR	NR	NR	67
Liang et al. (2021)[68]	China	YOLOv3 SSAM (CNN)	Central South University (own dataset)	Retrospective	12,909	NR	NR	6537	96.6	85.7	89.12	95.45	NR
Ke et al. (2021)[69]	China	CNN	Shanxi Tumor Hospital	Retrospective	130	NR	NR	NR	NR	NR	NR	NR	85
Zhu et al. (2021)[70]	China	AIATBS (CNN)	Multicenter prospective samples	Prospective + Retrospective	9215	NR	NR	NR	83.78	94.54	87.86	92.51	NR
Ma et al. (2021)[87]	China	FPN (CNN)	Alibaba Tianchi competition	Retrospective	4107	2792	493	822	95	NR	NR	NR	NR
Cheng et al. (2021)[71]	China	RNN	Maternal and Child Hospital of Hubei Province (Multi-center)	Retrospective	3545	8	1	1	92.8	95.3	86.86	97.55	NR
Kanavati et al. (2022)[72]	Japan	CNN + RNN (CNN)	A private clinical laboratory	Retrospective	1953	1503	150	300	85	91.1	40.48	98.84	NR
Shinde et al. (2022)[73]	India	DeepCyto (CNN)	Guwahati and Guwahati Medical College and Hospital	Retrospective	963	NR	NR	193	100	100	100	100	NR
Xu et al. (2022)[74]	China	Faster R-CNN	NR	Retrospective	NR	6666	NR	744	87.7	NR	NR	NR	NR
Wang et al. (2022)[56]	China	3cDe-Net (CNN)	Alibaba Cloud TianChi Company	Retrospective	13,254	8	1	1	99.4	94.86	95.09	99.37	NR
Alsalatie et al. (2022)[75]	India	EfficientNet-B3 (CNN)	Babina Diagnostic Pvt. Ltd + Clinical + hospital	Retrospective	963	NR	NR	NR	100	100	100	100	NR
Chowdary et al. (2023)[58]	China	SE-UNet (CNN)	Sapporo Medical University	Retrospective	1780	1650	330	130	97.18	NR	NR	NR	NR
Nambu et al. (2022)[76]	Japan	YOLOv4 + ResNeSt (CNN)	ISBI 2014 dataset	Retrospective	919	575	122	222	100	49	70.52	100	NR
Du et al. (2023)[77]	China	ResNet (CNN)	Two HOSPITAL	Retrospective	109,309	NR	NR	10,929	94.19	81.49	93.73	82.68	NR
Hamdi et al. (2023)[78]	America	RF-ResNet50-VGC19 (CNN)	(CESC) dataset	Retrospective	962	8	2	2	97.4	99.2	98.57	98.38	98.75
Kurita et al. (2023) 79	Japan	EfficientNet-B3 (CNN)	JA Shizuoka Kohseiren Enshu Hospital	Retrospective	39,990	25,559	5481	8950	72.5	92.6	78.06	90.26	90.8
Xue et al. (2023)[80]	China	CITL-AI (CNN)	Nine hospitals	Retrospective	3514	NR	NR	NR	87.2	91.5	35.54	99.25	89.3
Chantziantoniou et al. (2023)[81]	America	BestCyte Cell Sorter Imaging System (CNN)	Courtesy of Marlboro-Chesterfield Pathology	Retrospective	NR	NR	NR	NR	95.75	97.51	97.64	95.53	NR
Bai et al. (2024)[91]	China	ResNet50 (DNN)	The Third Affiliated Hospital of Zhengzhou University	Retrospective	NR	89,435	NR	1156	94.88	31.34	77.21	71.43	NR

(Table 2 continues on next page)

Study	Country	Algorithm	Dataset	Study design	No. of images	Training size	Validation size	Testing size	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC
(Continued from previous page)													
Civit-Masot et al. (2024)[92]	Multi-country	CNN	Mendeley liquid cytology dataset	Retrospective	300	210	30	60	100	100	100	100	NR
Wang et al. (2024)[93]	Multi-country	ResNet50 + random forest	Three institutions (America + China)	Prospective	16,056	9316	6132	608	100.00	98.90	88.24	100	99.5
Yang et al. (2024)[94]	China	FPN (CNN)	The First Medical Center of the Chinese PLA General Hospital	Retrospective	NR	NR	NR	1231	88.2	26.3	60.2	63.8	57.2
Zeng et al. (2024)[95]	China	AICyte	Four different hospital systems in China	Retrospective	NR	NR	NR	163,848	90.77	49.65	10.55	98.8	85.39

NR, not report; Se, sensitivity; Sp, specificity.

Table 2: **Characteristics of included studies about TCT.**

the pooled sensitivity was 97% (95% CI 95–99; I^2 = 99.9%), the pooled specificity was 94% (95% CI 85–98; I^2 = 100%), the pooled PPV was 84% (95% CI 64–98; I^2 = 100%), and the pooled NPV was 96% (95% CI 94–98; I^2 = 99.3%) (Figs. 4 and 5; Supplementary Table S2). Additionally, the areas under the summary receiver operating characteristic curve (AUC) of AI-assisted Pap smears and TCT for cervical cytology screening were 0.99 (95% CI 0.97–0.99) and 0.99 (95% CI 0.98–1.00), respectively (Supplementary Figs. S1 and S2).

We also conducted a subgroup analysis to evaluate the performance of AI-assisted cervical cytology in both developed and developing countries. With respect to AI-assisted Pap smears, there were no significant differences in the sensitivity and specificity between developed and developing countries, with values of 96% (95% CI 86–99; I^2 = 99.9%) vs. 94% (95% CI 92–96; I^2 = 83.8%; P = 0.53), 95% (95% CI 86–98; I^2 = 99.9%) vs. 94% (95% CI 85–98; I^2 = 99.6%; P = 0.51). The accuracy, PPV, and NPV significantly differed between the two groups, with values of 94% (95% CI 89–99; I^2 = 99.8%) vs. 93% (95% CI 91–96; I^2 = 99.3%; P < 0·001), 94% (95% CI 83–99; I^2 = 99.1%) vs. 80% (95% CI 48–99; I^2 = 99.6%; P < 0.001), and 92% (95% CI 80–99; I^2 = 99.1%) vs. 98% (95% CI 97–99; I^2 = 76.8%; P = 0.014), respectively (Supplementary Table S2). In AI-assisted TCT, there were significant differences in the accuracy, sensitivity, specificity, PPV, or NPV between developed and developing countries. The values were as follows: accuracy, 94% (95% CI 92–96; I^2 = 99.5%) vs. 87% (95% CI 78–96; I^2 = 100%; P < 0.001); sensitivity, 98% (95% CI 90–100; I^2 = 99.8%) vs. 96% (95% CI 93–98; I^2 = 99.7%; P = 0.045); specificity, 98% (95% CI 87–100; I^2 = 99.9%) vs. 87% (95% CI 71–94; I^2 = 100%; P < 0.001); PPV, 89% (95% CI 75–98; I^2 = 99.0%) vs. 81% (95% CI 53–98; I^2 = 100%; P < 0.001); and NPV, 98% (95% CI 93–100; I^2 = 99.3%) vs. 94% (95% CI 91–97; I^2 = 99.4%; P < 0.001) (Supplementary Table S2).

For the performance of AI and experienced colposcopists in diagnosing cervical lesions via colposcopy, the pooled accuracy of AI was 81% (95% CI 77–84; I^2 = 98.9), whereas the pooled accuracy of experienced colposcopists was 74% (95% CI 69–79; I^2 = 97.3) (Figs. 6 and 8; Table 3). The pooled sensitivity of AI was 86% (95% CI 76–92; I^2 = 97.3) and its pooled specificity was 83% (95% CI 73–90; I^2 = 97.8) (Fig. 7; Table 3). In comparison, the pooled sensitivity of experienced colposcopists was 85% (95% CI 71–93; I^2 = 96.4) and their pooled specificity was 67% (95% CI 46–83; I^2 = 97.1) (Fig. 9; Table 3). The pooled PPV and NPV of AI were 82% (95% CI 74–89; I^2 = 98.6) and 80% (95% CI 69–89; I^2 = 98.9), respectively, whereas those of experienced colposcopists were 76% (95% CI 60–89; I^2 = 99.2) and 75% (95% CI 58–89; I^2 = 99.0) (Table 3). Based on

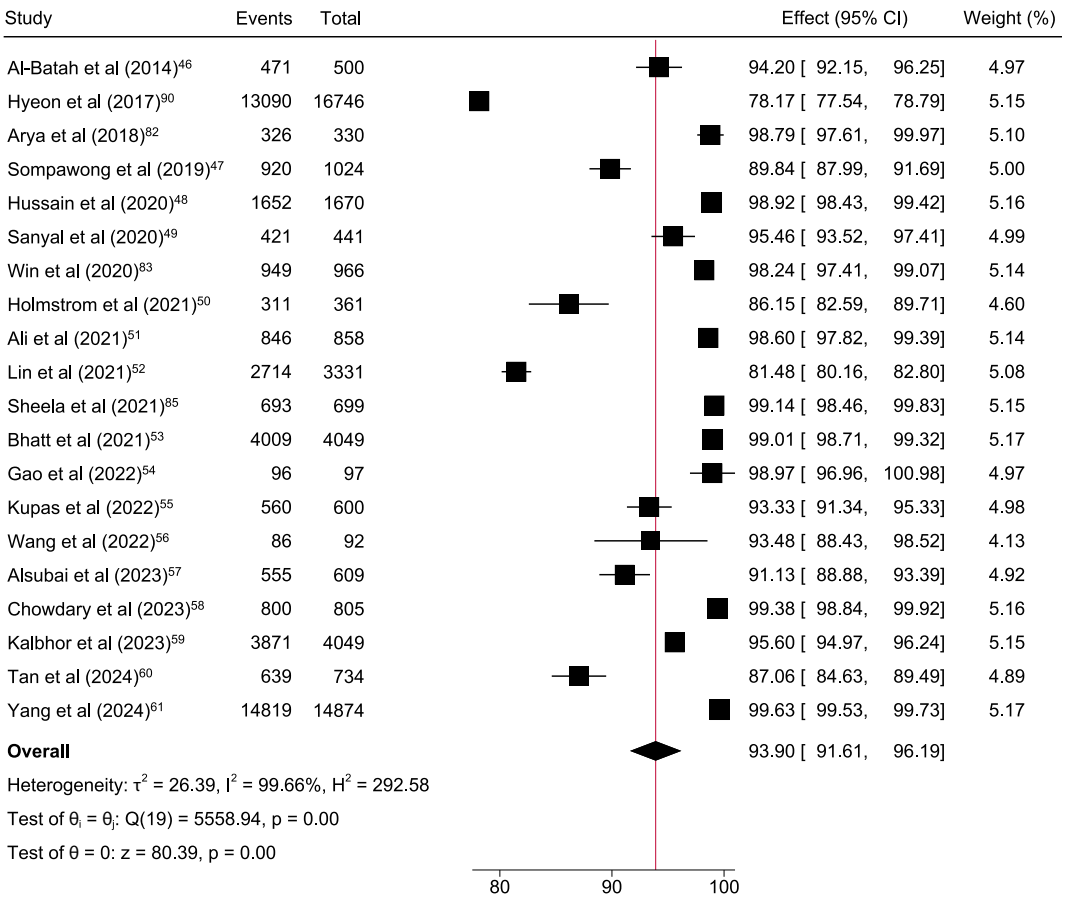

Fig. 2: Diagnostic accuracy of AI in Papanicolaou (Pap) smears. Meta-analysis of the diagnostic accuracy of Pap smears in the 20 studies included. Solid vertical lines show the pooled estimates. AI, artificial intelligence; CI, confidence interval.

pooled data analysis, AI demonstrated superior accuracy in colposcopic examinations compared with clinicians. A similar result was achieved through the application of the random-effects model (OR 1.75; 95% CI 1.33–2.31; P < 0.0001; I^2 = 93%) (Fig. 10). Furthermore, the AUCs of AI-assisted and clinician-performed (experienced colposcopists) colposcopic examinations were 0.91 (95% CI 0.88–0.93) and 0.85 (95% CI 0.81–0.88), respectively (Supplementary Figs. S3 and S4).

A subgroup analysis was subsequently conducted to compare the performance of AI and experienced colposcopists in the colposcopic diagnosis of cervical lesions between developed and developing countries. For AI, only specificity was not statistically significant, with values for developed and developing countries of 83% (95% CI 68–92; I^2 = 97.5) and 82% (95% CI 69–90; I^2 = 98.1; P = 0.321), respectively. Other diagnostic indicators, including the accuracy, sensitivity, PPV, and NPV, were significantly different between developed and developing countries (80% (95% CI 74–86; I^2 = 97.6) vs. 82% (95% CI 77–86; I^2 = 98.3; P < 0.05), 89% (95% CI 75–96; I^2 = 96.8) vs. 82% (95% CI 72–88; I^2 = 97.3; P < 0.05), 87% (95% CI 79–93; I^2 = 94.1) vs. 77% (95% CI 64–88; I^2 = 99.2; P < 0.05), and 78% (95% CI 58–92; I^2 = 97.9) vs. 82% (95% CI 72–90; I^2 = 98.7; P < 0.05), respectively) (Supplementary Table S3). For experienced colposcopists, the accuracy was not statistically significant, with values for developed and developing countries of 74% (95% CI 68–80; I^2 = 85.7) and 74% (95% CI 67–81; I^2 = 98.6; P = 0.51), respectively. The sensitivity, specificity, PPV, and NPV for developed and developing countries were 89% (95% CI 73–96; I^2 = 85.1) vs. 74% (95% CI 63–83; I^2 = 97.3; P < 0.05), 68% (95% CI 35–89; I^2 = 96.2) vs. 75% (95% CI 48–94; I^2 = 99.7; P < 0.05), 83% (95% CI 66–95; I^2 = 96.2) vs. 65% (95% CI 39–87; I^2 = 99.5; P < 0.05), and 71% (95% CI 56–85; I^2 = 87.8) vs. 83% (95% CI 79–86; I^2 = 90.3; P < 0.05), respectively, with statistically significant differences (Supplementary Table S4). Moreover, AI achieved greater accuracy in colposcopic examinations than did clinicians in the subgroup analysis of developed countries (OR 0.08; 95% CI 0.03–0.13; P < 0.001; I^2 = 53%) and developing countries (OR 0.08; 95% CI 0.02–0.14; P < 0.05; I^2 = 97%) (Supplementary Fig. S5).

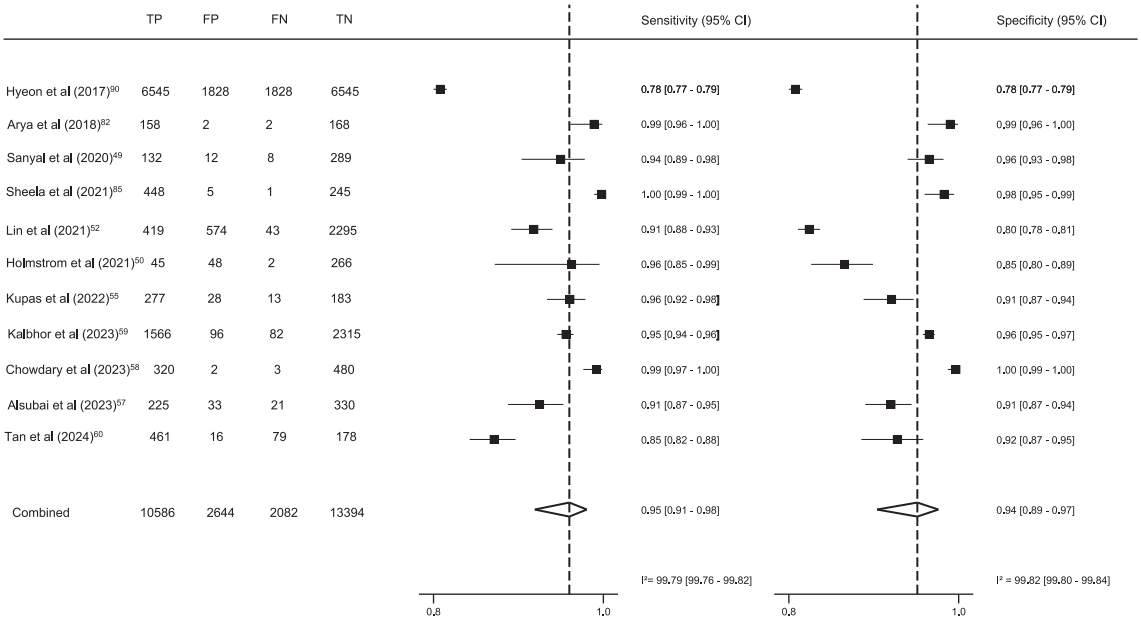

Fig. 3: Forest plot demonstrating diagnostic sensitivity and specificity of AI in Pap smears. Meta-analysis was performed in the 11 studies included. Dashed vertical lines show the pooled estimates. TP, true positive; FN, false negative; FP, false positive; TN, true negative.

Given that colposcopic diagnosis of HSIL or worse (HSIL+) requiring immediate treatment is crucial for the prevention of cervical cancer, subgroup analysis was conducted to compare the ability of AI and experienced colposcopists to detect LSIL or worse (LSIL+) and HSIL+. For AI-assisted colposcopy, none of the diagnostic indicators were significantly different in detecting LSIL+ and HSIL+. The accuracy, sensitivity, specificity, PPV, and NPV for detecting LSIL+ and HSIL+ were as follows: 84% (95% CI 76–92; $I^2 = 99.5$; $P = 0.87$) vs. 81% (95% CI 76–87; $I^2 = 99.0$), 86% (95% CI 75–92; $I^2 = 98.9$; $P = 0.65$) vs. 89% (95% CI 77–95; $I^2 = 98.2$), 85% (95% CI 75–92; $I^2 = 99.6$; $P = 0.37$) vs. 86% (95% CI 74–93; $I^2 = 98.3$), 82% (95% CI 65–94; $I^2 = 99.3$; $P = 0.341$) vs. 85% (95% CI 73–93; $I^2 = 98.6$), and 86% (95% CI 79–92; $I^2 = 96.8$; $P = 0.19$) vs. 82% (95% CI 64–94; $I^2 = 99.3$) (Supplementary Table S3). For experienced colposcopists, both the sensitivity, PPV, and NPV were statistically significant, with the following values for detecting LSIL+ and HSIL+: 70% (95% CI 68–72; $P < 0.05$) vs. 87% (95% CI 74–94; $I^2 = 96.7$), 61% (95% CI 59–63; $P < 0.05$) vs. 80% (95% CI 62–94; $I^2 = 98.8$), and 80% (95% CI 79–82; $P < 0.05$) vs. 73% (95% CI 59–86; $I^2 = 96.4$). However, the accuracy and specificity for detecting LSIL+ and HSIL+ were 75% (95% CI 68–83; $I^2 = 97.7$; $P = 0.87$) vs. 74% (95% CI 69–80; $I^2 = 95.7$) and 73% (95% CI 71–74; $P = 0.581$) vs. 71% (95% CI 44–88; $I^2 = 99.0$), respectively, with no significant differences (Supplementary Table S4). Additionally, AI outperformed clinicians in diagnosing LSIL+ patients (OR 1.84; 95% CI 1.28–2.65; $P < 0.01$; $I^2 = 89\%$) and HSIL+ patients (OR 1.43; 95% CI 1.07–1.91; $P < 0.05$; $I^2 = 86\%$) (Supplementary Fig. S6).

Fagan's nomogram indicates that if the prevalence of cervical lesions is 20%, the true positive rate for AI-assisted Pap smears, TCT, and colposcopy is 81%, 80%, and 56%, respectively, with corresponding probabilities of having cervical lesions if the diagnosis is negative being 1%, 1%, and 4%. For expert colposcopists, the true positive rate is 39% and the false negative rate is 5% (Supplementary Figs. S7–S10). The Deeks' funnel plot for the 60 studies included in our meta-analysis indicated that there was no evidence of publication bias for AI-assisted colposcopy or clinician-performed colposcopy. However, there was publication bias in the results of AI-assisted Pap smears and AI-assisted TCT (Supplementary Figs. S11–S14). We used the leave-one-out method for sensitivity analysis of the results. Studies were removed one at a time, and subsequent meta-analyses were conducted to observe any changes in the effect size. The results of the subsequent meta-analyses were similar to the previous combined effect size results after each study was removed. Furthermore, we applied the trim-and-fill method to assess and adjust for publication bias in the random-effects model. The results indicate that potential publication bias had a minimal impact on the model (Supplementary Figs. S15 and S16). In AI-based diagnosis, regression analysis of the results revealed that the heterogeneity mainly originated from the sample size and ground truth. For AI-assisted TCT, the results indicated that the heterogeneity was primarily due to the sample size, subject, and predesign. For AI-assisted pap smear, the heterogeneity was due mainly to the

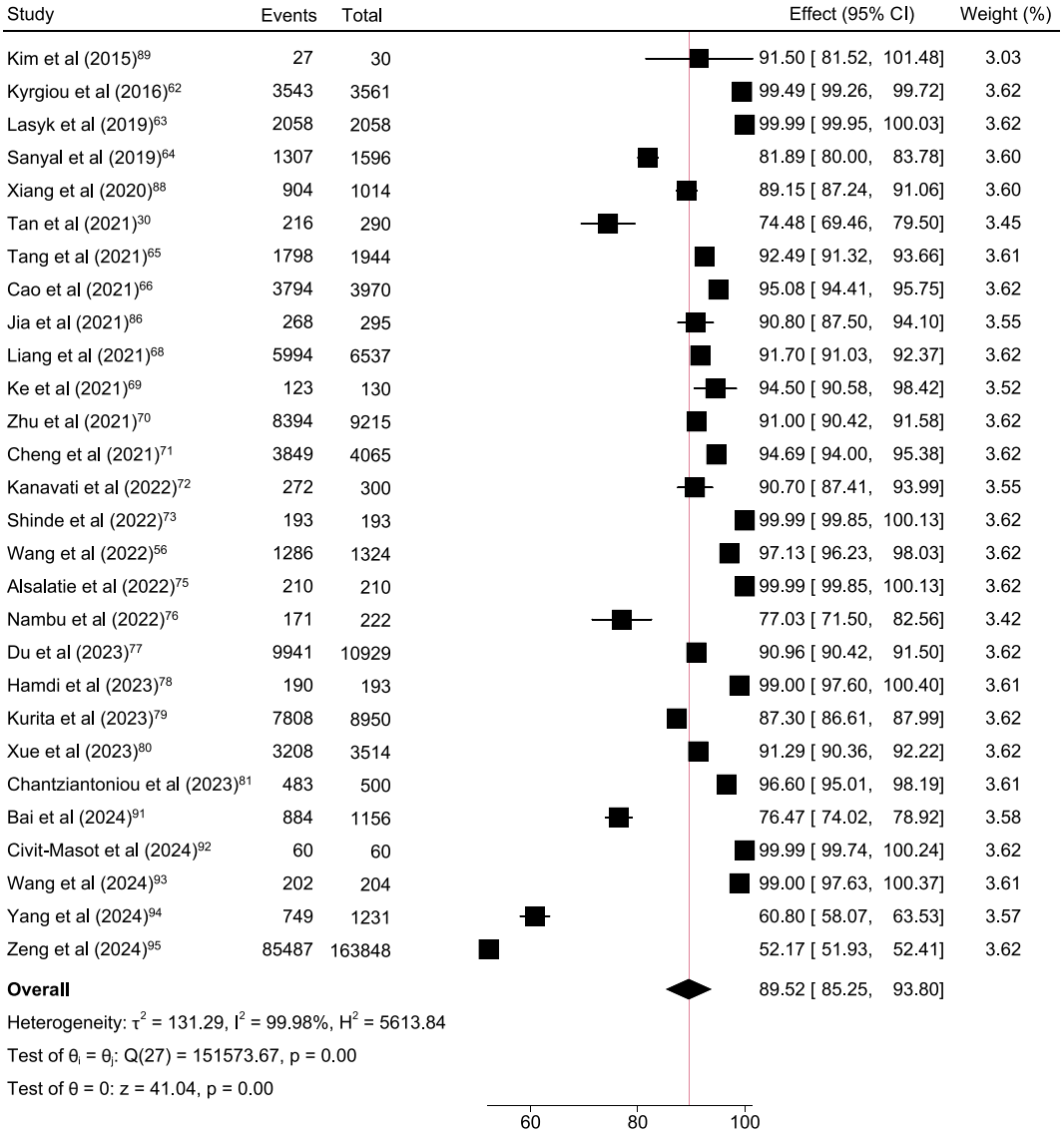

Fig. 4: Forest plot demonstrating diagnostic accuracy of AI in TCT. Meta-analysis of AI diagnostic accuracy in TCT based on 28 included studies, with solid vertical lines representing the pooled estimates. TCT, ThinPrep cytologic test.

intervention type and sample size (Supplementary Figs. S21–S23). Although we conducted subgroup analyses for various diagnostic tests, the heterogeneity remained substantial. The bagplot diagrams for the cytology and colposcopy diagnostic tests revealed that several data points in each study fell outside the 95% CI. After these studies were excluded and a sensitivity analysis was performed, the heterogeneity did not change significantly (Supplementary Figs. S24–S27). Each study was assessed for quality via the QUADAS-2 tool. The primary risk of bias was identified in the selection of cases. None of the studies presented a high risk of bias in more than three domains (Supplementary Figs. S28–S33). Additionally, the evidence was evaluated using the GRADE system (Supplementary Tables S5–S7), and the overall quality was rated as moderate.

Discussion

To our knowledge, this is the first systematic review to assess the diagnostic accuracy, sensitivity, and specificity of AI-assisted cervical cytology screening and colposcopy for detecting CIN and cervical cancer. Our main findings indicate that AI-assisted cervical cytology screening can distinguish between normal and abnormal cytological results with high accuracy, sensitivity, and specificity (Pap smear: 94%, 95%, and

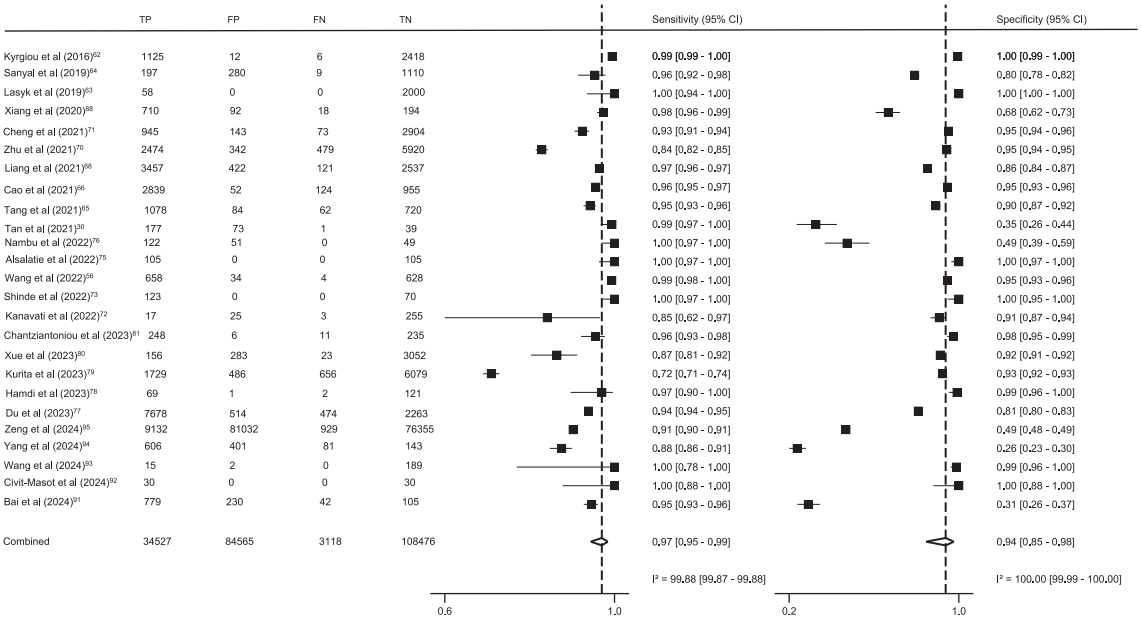

Fig. 5: The diagnostic sensitivity and specificity of AI in TCT. A meta-analysis was conducted on the 25 included studies, with dashed vertical lines representing the pooled estimates. TP, true positive; FN, false negative; FP, false positive; TN, true negative.

94%; TCT: 90%, 97%, 94%, respectively). Additionally, AI demonstrated a substantial agreement rate with pathological results of 81%, which is considered the gold standard for grading all colposcopic impressions, surpassing that of experienced colposcopists (74%; OR 1.75; 95% CI 1.33–2.31). Furthermore, AI exhibited superior performance over experienced colposcopists in diagnosing LSIL+ and HSIL+ patients. Subgroup analysis revealed that the diagnostic accuracy of AI exceeded that of experienced colposcopists in both developing and developed countries. These findings underscore the significant advantages of AI in cervical cancer screening and diagnosis, particularly in colposcopy and cervical cytology screening. Compared to clinical doctors, AI not only had greater overall accuracy but also exhibited higher sensitivity and specificity in diagnosing various lesions. These results highlight the substantial potential and application value of AI in the prevention and early diagnosis of cervical cancer.

The accuracy of cervical cytology screening is often suboptimal, as traditional screening methods exhibit considerable variability in sensitivity and specificity, resulting in false-negative or false-positive outcomes.[80] This inaccuracy increases the risk of missed or incorrect diagnoses, thereby affecting the timing of patient treatment and prognosis of disease. Furthermore, traditional screening relies on experienced cytopathologists for result interpretation; each smear contains 20,000–50,000 cells, making the process time-consuming and heavily dependent on individual expertise and skill. Cytopathologists with less experience may overlook high-risk lesions, impacting the effectiveness of screening. In resource-limited regions, particularly in developing countries, there is a lack of trained personnel and equipment to conduct effective cervical cytology screening. This disparity in resources leads to low screening coverage, preventing timely detection and treatment in many high-risk patients. Our meta-analysis revealed that AI achieved high accuracy in cancer screening by detecting and classifying abnormal cells through the analysis of many cell images. The accuracy for Pap smears was 94%, and that for TCT was 90%. AI, being 380 times faster than typical pathologists, accelerates the screening process, which is particularly vital in resource-limited areas. AI assists cytopathologists in smear analysis, thereby alleviating their workload by efficiently processing large volumes of data and allowing them to focus on more complex cases. Furthermore, AI helps minimize regional disparities by providing high-quality screening services through remote diagnosis and intelligent analysis, thus narrowing the healthcare gap between urban and rural regions.

Currently, the primary techniques for diagnosing CIN and cervical cancer are colposcopy and guided biopsy.[39] The effectiveness of colposcopy is limited by a strong reliance on the subjective experience of the operator, significant inter- and intraoperator variability, and a shortage of skilled colposcopists.[41] Additionally, the complexity of comprehensive colposcopy training programs and the need for standardized diagnostic

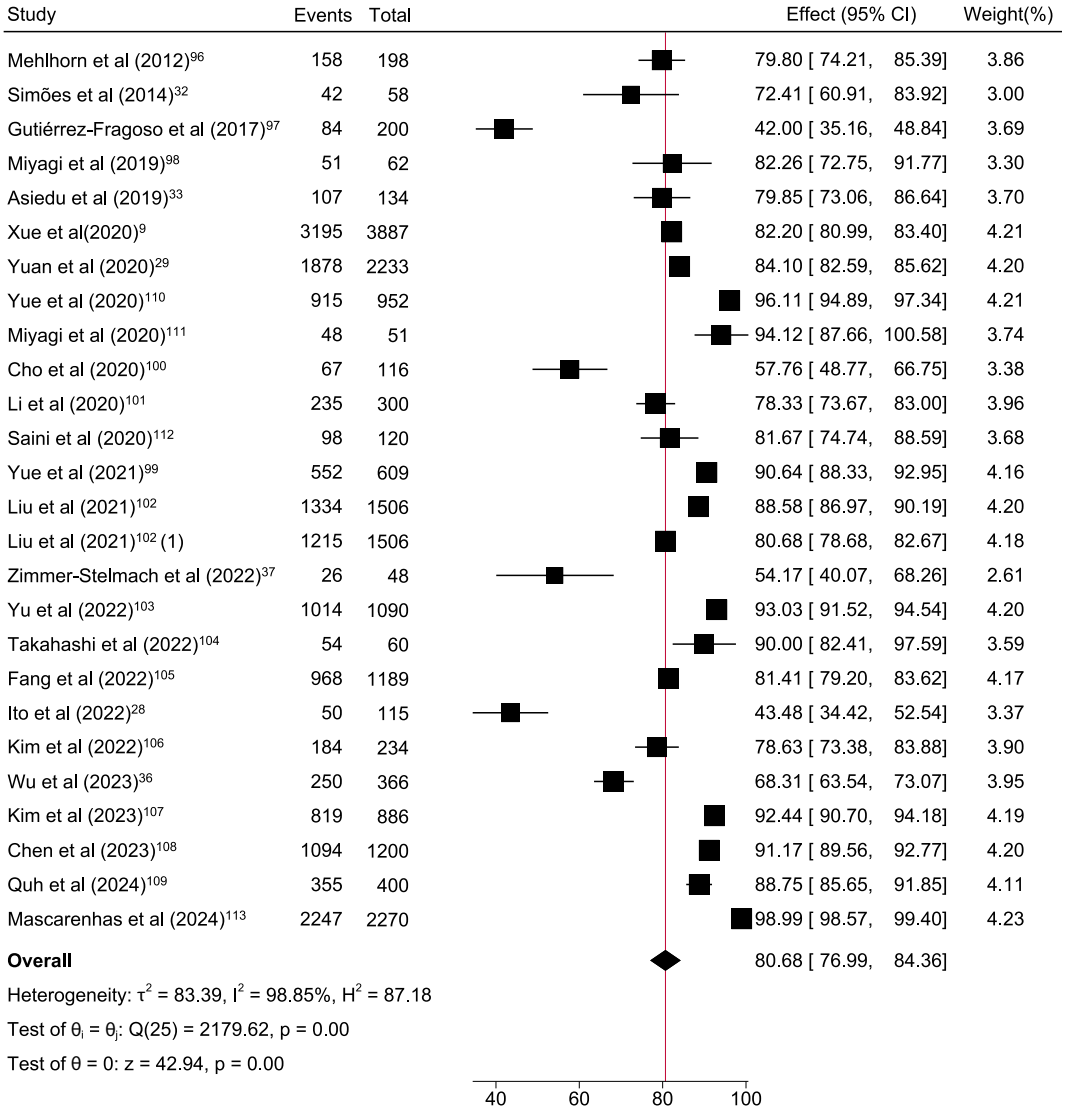

Fig. 6: Forest plot illustrating the diagnostic accuracy of AI in colposcopy. A meta-analysis was conducted on the diagnostic accuracy across the 25 included studies, with solid vertical lines indicating the pooled estimates.

criteria and rigorous quality control are difficult to achieve consistently, especially for colposcopists with limited diagnostic skills. This variability can lead to inconsistent reporting and documentation of colposcopy findings. Structured colposcopy training programs are essential for improving the diagnostic capabilities of colposcopists. However, the practical implementation of these programs to enhance diagnostic performance in a short timeframe is challenging, especially in LMICs.[41] Fortunately, numerous investigations have shown that AI has the potential to mitigate these challenges.[9,29,36,39,41] In the present meta-analysis, AI demonstrated superior accuracy in colposcopic examinations compared with clinicians (OR 1.75; 95% CI 1.33–2.31). Furthermore, the AUCs of AI-assisted and colposcopist-performed colposcopic examination were 0.91 and 0.85, respectively. Moreover, AI achieved greater accuracy in colposcopic examinations than did clinicians in the subgroup analysis of developed countries (OR 0.08; 95% CI 0.03–0.13) and developing countries (OR 0.08; 95% CI 0.02–0.14). AI can assist colposcopists by providing more accurate colposcopy image interpretations, detecting underlying CIN, and guiding biopsy site selection. The automation of the colposcopy examination process could establish a novel cervical cancer screening model, reduce the incidence of false negatives and false positives, and improve the accuracy of colposcopy-based diagnoses and cervical biopsies.

Parameters	AI	Clinician
Accuracy	81% (77–84; 98.9%) 26 datasets	74% (69–79; 97.3%) 14 datasets (P < 0.05)
Sensitivity	86% (76–92; 97.3%) 16 datasets	85% (71–93; 96.4%) 8 datasets (P < 0.05)
Specificity	83% (73–90; 97.8%) 16 datasets	67% (46–83; 97.1%) 8 datasets (P < 0.05)
Positive predictive value	82% (74–89; 98.6%) 16 datasets	76% (60–89; 99.2%) 8 datasets (P < 0.05)
Negative predictive value	80% (69–89; 98.9%) 16 datasets	75% (58–89; 99.0%) 8 datasets (P < 0.05)

Clinician, experienced colposcopist. Values in parentheses are 95% confidence interval; I^2.

Table 3: **Summary of pooled rates for colposcopy.**

To date, there are no established guidelines for the application of AI-assisted cervical cytology screening and colposcopy. Numerous previous studies have demonstrated that AI has acceptable performance in both cervical cytology screening and colposcopy. However, no relevant meta-analysis has been conducted until now. In our meta-analysis, we found that AI achieved a high accuracy rate in cervical cytology screening. Additionally, AI outperformed experienced colposcopists in terms of diagnostic accuracy. Given that colposcopy relies heavily on subjective experience, and considering the shortage of experienced cytopathologists and colposcopists in developing countries, as well as the lack of relevant training and quality control, the diagnostic accuracy of cervical biopsy in detecting CIN is reported to be relatively low, ranging from 30% to 70%, especially in LMICs. This challenge could hinder the achievement of the global goal of eliminating cervical cancer by 2030. Policymakers should recognize the importance of AI in bridging the healthcare gap between developed and developing countries and its significant role in improving cervical cancer screening processes in developing nations.

In our meta-analysis, we identified varying levels of heterogeneity and potential bias among the included studies, necessitating careful interpretation of the results. The Deeks' funnel plot indicated no evidence of publication bias for AI-assisted colposcopy or clinician-performed colposcopy; however, publication bias was observed in AI-assisted Pap smears and AI-assisted TCT. To evaluate the robustness of our findings, we employed the leave-one-out method for sensitivity analysis, confirming that no single study significantly influenced the overall results. The application of the trim-and-fill method within the random-effects model indicated a minimal impact from publication bias, further reinforcing our conclusions. Heterogeneity primarily arose from factors such as sample size and the definition of ground truth in AI-assisted TCT, while intervention type contributed to heterogeneity in AI-assisted colposcopy. Quality assessment using the QUADAS-2 tool revealed risks associated with case selection; however, no studies exhibited a high risk of bias across multiple domains. Additionally, the evidence was evaluated using the GRADE system, resulting in an

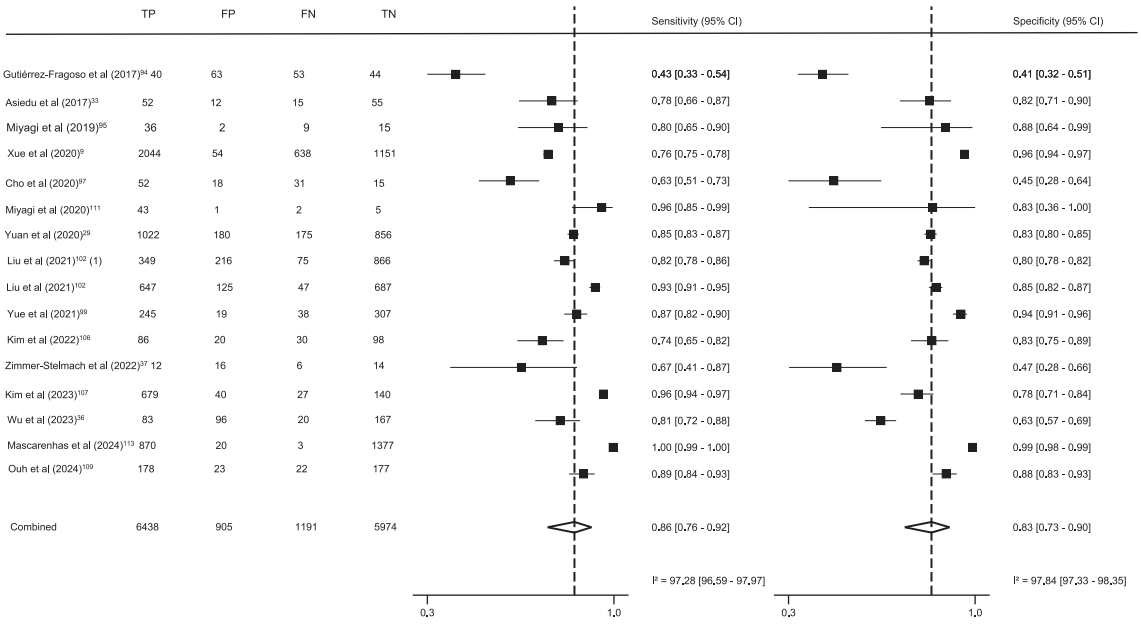

Fig. 7: The diagnostic sensitivity and specificity of AI in colposcopy were assessed through a meta-analysis of 15 included studies. Dashed vertical lines indicate the pooled estimates. TP, true positive; FN, false negative; FP, false positive; TN, true negative.

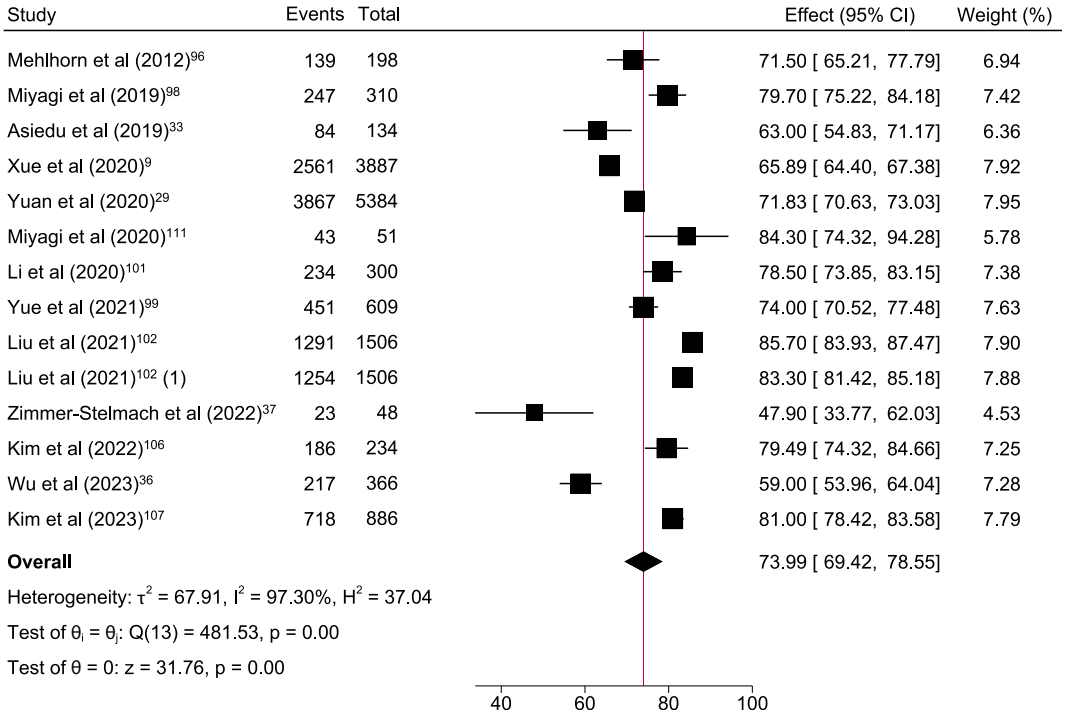

Fig. 8: Diagnostic accuracy of clinicians in colposcopy. A meta-analysis of 13 included studies assessed clinicians' diagnostic accuracy in colposcopy, with solid vertical lines indicating the pooled estimates. Clinicians, experienced colposcopists.

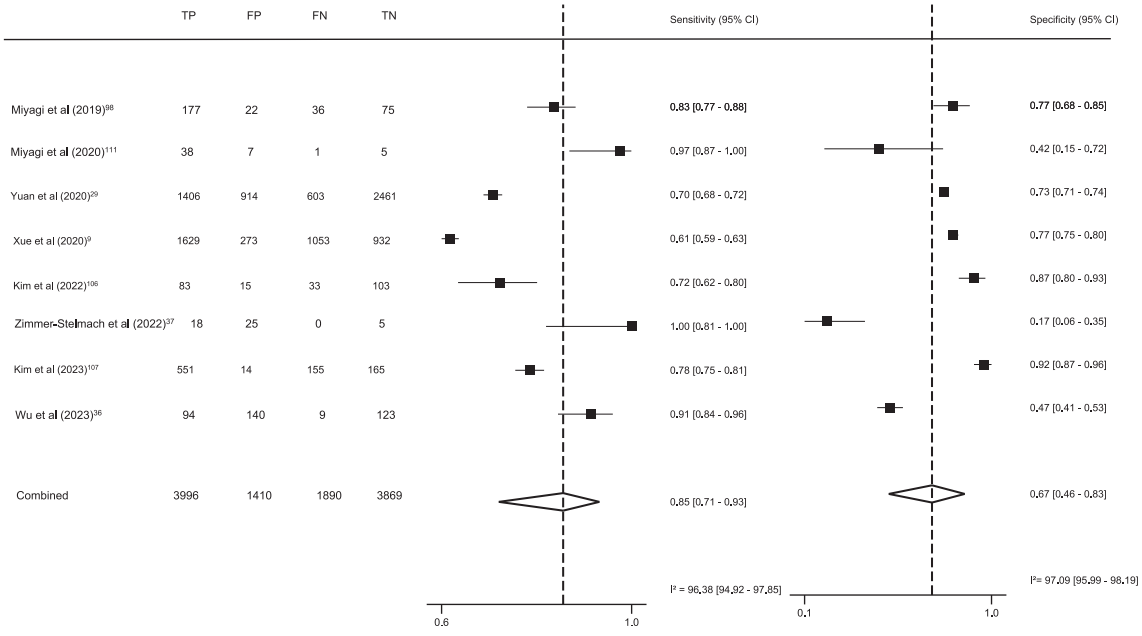

Fig. 9: The diagnostic sensitivity and specificity of clinicians in colposcopy were analyzed in a meta-analysis of eight included studies, with dashed vertical lines marking the pooled estimates. TP, true positive; FN, false negative; FP, false positive; TN, true negative; Clinicians, experienced colposcopists.

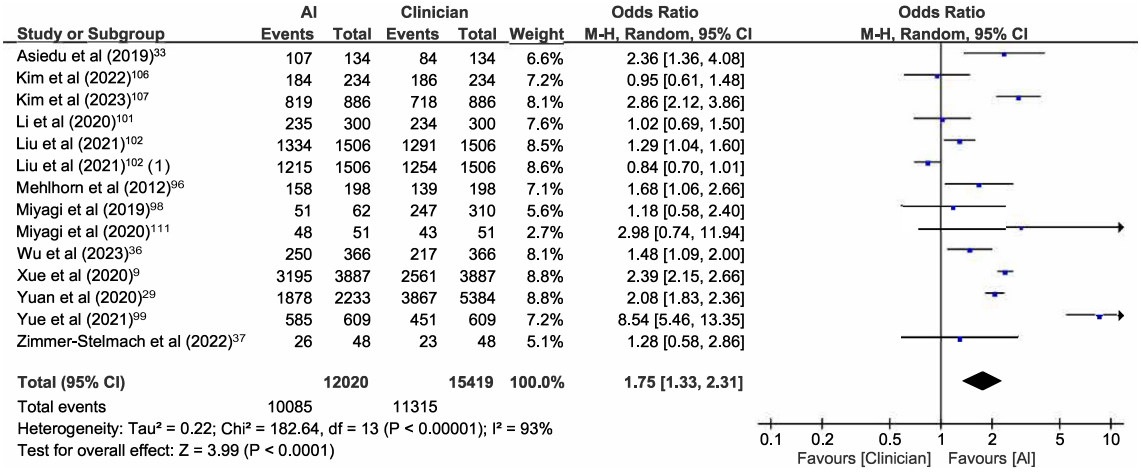

Fig. 10: A random-effects forest plot illustrating the comparative diagnostic accuracy of AI models vs. clinicians in colposcopy. This plot highlights the differences in performance between AI-based models and human clinicians in accurately diagnosing conditions through colposcopy. OR, odds ratio; Clinicians, experienced colposcopists; AI, artificial intelligence.

overall quality rating of moderate. Overall, addressing heterogeneity and bias is crucial for ensuring the reliability and applicability of our findings in real-world clinical settings.

This is the first systematic review and meta-analysis to comprehensively evaluate the performance of AI-assisted cervical cytology screening and colposcopy. This study incorporated a large integrated sample size, covered various AI algorithms, and compared the performance of AI-assisted colposcopy with that of clinician-performed colposcopy. It also analyzed the ability of AI to grade LSIL+ and HSIL+ and compared the diagnostic efficiency of AI across countries with different development levels. However, several limitations exist. First, the majority of studies utilized small sample sizes, retrospective data, and single-center datasets to validate AI systems, and prospective studies and external validation are lacking. Second, owing to the small sample sizes, most studies were underpowered in terms of the primary outcomes. Some algorithms used in these studies are highly unstable, meaning that slight changes in the data can significantly alter the decision-making process. Moreover, there is a concern regarding overfitting, where models perform exceptionally well on training data but fail to generalize effectively to new, unseen data, undermining their real-world applicability. Third, the datasets used may primarily originate from specific regions or certain types of healthcare institutions, lacking broad representation across different races, ages, and geographic areas. Fourth, the standards and annotation methods used in different studies are not uniform, potentially leading to poor comparability of results. Fifth, some studies may not have provided detailed descriptions of the technical specifics and parameter settings used, making the results difficult to replicate and verify. Finally, these trials lacked statistical analysis of the cost-effectiveness and diagnostic efficiency of AI.

Despite the superior performance of AI-assisted cervical cytology screening and colposcopy, several challenges need to be addressed in future research. First, more multicentre, large-sample prospective studies are needed to internally and externally validate AI performance. Second, AI algorithms typically require millions of observations to achieve acceptable performance levels. Future research should focus on the standardization, authenticity, and accuracy of data collection. Third, given the diversity of AI algorithms, improving the compatibility of AI software is crucial. Fourth, further enhancing the diagnostic performance of AI to reduce misdiagnosis and missed diagnosis rates is essential. Additionally, varying prevalence rates of abnormalities across different populations can significantly affect the utility and predictive value of AI models, making it important to contextualize AI performance in low-prevalence settings. Moreover, enriched datasets may lead to inflated performance metrics, highlighting the importance of external validation for assessing model robustness in real-world screening settings. Addressing the challenges of data scarcity and model validation will ensure the reliability and practicality of AI in diagnosing rare cases. Fifth, to ensure the ethical, legal, and effective clinical application of AI in medicine, it is crucial to develop regulations, address ethical concerns such as data privacy and algorithm fairness, and establish AI image cloud platforms to enhance diagnostic support, improve clinician skills, and expand healthcare access in resource-limited areas. Finally, improving research on health economics and the diagnostic efficiency of AI is essential.

AI demonstrates high accuracy, sensitivity, and specificity in cervical cytology screening and colposcopy. Notably, AI has superior accuracy in colposcopic examinations compared with experienced colposcopists. The analysis of data from both developed and developing countries highlights the potential clinical significance of AI in improving cervical cancer screening performance in LMICs, thereby accelerating the elimination of cervical cancer worldwide. We urge clinicians and public health program designers to be aware of the remarkable diagnostic accuracy of AI and its ability to assist healthcare professionals, especially in LMICs with limited healthcare resources.

Contributors

All authors have made substantial contributions to this study. RX and HXX conceptualised and designed the study. LL wrote the original manuscript. LL, QS and YNC collected the data. LL, JGL, QS, and YNC did the statistical analyses. LL, JGL, QS, and RX accessed and verified the underlying data. HXX offered valuable suggestions and consultations on study design and data analyses. JGL provided critical revisions and were responsible for reviewing and editing the manuscript. All authors reviewed and edited the manuscript and approved the final version. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data sharing statement

This systematic review and meta-analysis used data that were available in previously published studies.

Declaration of interests

No potential conflicts of interest relevant to this article were reported.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant No. 81901493) and the Shanghai Pujiang Program (Grant No. 21PJD006).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102992.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
2 Castle PE, Einstein MH, Sahasrabuddhe VV. Cervical cancer prevention and control in women living with human immunodeficiency virus. *CA Cancer J Clin*. 2021;71(6):505–526.
3 Das M. WHO launches strategy to accelerate elimination of cervical cancer. *Lancet Oncol*. 2021;22(1):20–21.
4 Simelela PN. WHO global strategy to eliminate cervical cancer as a public health problem: an opportunity to make it a disease of the past. *Int J Gynaecol Obstet*. 2021;152(1):1–3.
5 Kuehn BM. WHO launches global push to eliminate cervical cancer. *JAMA*. 2021;325(3):213.
6 Brisson M, Kim JJ, Canfell K, et al. Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet*. 2020;395(10224):575–590.
7 de Sanjose S, Holme F. What is needed now for successful scale-up of screening? *Papillomavirus Res*. 2019;7:173–175.
8 Perkins RB, Wentzensen N, Guido RS, et al. Cervical cancer screening: a review. *JAMA*. 2023;330(6):547–558.
9 Xue P, Tang C, Li Q, et al. Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. *BMC Med*. 2020;18(1):406.
10 Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*. 2000;132(10):810–819.
11 Schiffman M, Wentzensen N. Issues in optimising and standardising the accuracy and utility of the colposcopic examination in the HPV era. *Ecancermedicalscience*. 2015;9:530.
12 Brown BH, Tidy JA. The diagnostic accuracy of colposcopy - a review of research methodology and impact on the outcomes of quality assurance. *Eur J Obstet Gynecol Reprod Biol*. 2019;240:182–186.
13 Leeson SC, Alibegashvili T, Arbyn M, et al. The future role for colposcopy in Europe. *J Low Genit Tract Dis*. 2014;18(1):70–78.
14 Li J, Wang W, Yang P, et al. Analysis of the agreement between colposcopic impression and histopathological diagnosis of cervical biopsy in a single tertiary center of Chengdu. *Arch Gynecol Obstet*. 2021;304(4):1033–1041.
15 Muwonge R, Mbalawa CG, Keita N, et al. Performance of colposcopy in five sub-Saharan African countries. *BJOG*. 2009;116(6):829–837.
16 Kuo RYL, Harrison C, Curran TA, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology*. 2022;304(1):50–62.
17 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
18 Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3:118.
19 Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med*. 2022;28(1):31–38.
20 Zhou HY, Yu Y, Wang C, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat Biomed Eng*. 2023;7(6):743–755.
21 Aboutalib SS, Mohamed AA, Berg WA, et al. Deep learning to distinguish recalled but benign mammography images in breast cancer screening. *Clin Cancer Res*. 2018;24(23):5902–5909.
22 Cao K, Xia Y, Yao J, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat Med*. 2023;29(12):3033–3043.
23 Xi IL, Zhao Y, Wang R, et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. *Clin Cancer Res*. 2020;26(8):1944–1952.
24 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410.
25 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118.
26 Massad LS, Jeronimo J, Katki HA, et al. The accuracy of colposcopic grading for detection of high-grade cervical intraepithelial neoplasia. *J Low Genit Tract Dis*. 2009;13(3):137–144.
27 Yang B, Pretorius RG, Belinson JL, et al. False negative colposcopy is associated with thinner cervical intraepithelial neoplasia 2 and 3. *Gynecol Oncol*. 2008;110(1):32–36.
28 Ito Y, Miyoshi A, Ueda Y, et al. An artificial intelligence-assisted diagnostic system improves the accuracy of image diagnosis of uterine cervical lesions. *Mol Clin Oncol*. 2022;16(2):27.
29 Yuan C, Yao Y, Cheng B, et al. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Sci Rep*. 2020;10(1):11639.
30 Tan X, Li K, Zhang J, et al. Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study. *Cancer Cell Int*. 2021;21(1):35.
31 Bao H, Bi H, Zhang X, et al. Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: a multicenter, clinical-based, observational study. *Gynecol Oncol*. 2020;159(1):171–178.
32 Simões PW, Izumi NB, Casagrande RS, et al. Classification of images acquired with colposcopy using artificial neural networks. *Cancer Inform*. 2014;13:119–124.
33 Asiedu MN, Simhal A, Chaudhary U, et al. Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, pocket colposcope. *IEEE Trans Biomed Eng*. 2019;66(8):2306–2318.
34 Song D, Kim E, Huang X, et al. Multimodal entity coreference for cervical dysplasia diagnosis. *IEEE Trans Med Imaging*. 2015;34(1):229–245.

35 Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst*. 2019;111(9):923–932.
36 Wu A, Xue P, Abulizi G, et al. Artificial intelligence in colposcopic examination: a promising tool to assist junior colposcopists. *Front Med (Lausanne)*. 2023;10:1060451.
37 Zimmer-Stelmach A, Zak J, Pawlosek A, et al. The application of artificial intelligence-assisted colposcopy in a tertiary care hospital within a cervical pathology diagnostic unit. *Diagnostics (Basel)*. 2022;12(1):106.
38 Vargas-Cardona HD, Rodriguez-Lopez M, Arrivillaga M, et al. Artificial intelligence for cervical cancer screening: scoping review, 2009-2022. *Int J Gynaecol Obstet*. 2024;165(2):566–578.
39 Allahqoli L, Laganà AS, Mazidimoradi A, et al. Diagnosis of cervical cancer and pre-cancerous lesions by artificial intelligence: a systematic review. *Diagnostics (Basel)*. 2022;12(11):2771.
40 Hou X, Shen G, Zhou L, et al. Artificial intelligence in cervical cancer screening and diagnosis. *Front Oncol*. 2022;12:851367.
41 Xue P, Ng MTA, Qiao Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. *BMC Med*. 2020;18(1):169.
42 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(8):529–536.
43 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Contr Clin Trials*. 1986;7(3):177–188.
44 Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560.
45 Mohan BP, Adler DG. Heterogeneity in systematic review and meta-analysis: how to read between the numbers. *Gastrointest Endosc*. 2019;89(4):902–903.
46 Al-batah MS, Isa NA, Klaib MF, et al. Multiple adaptive neuro-fuzzy inference system with automatic features extraction algorithm for cervical cancer recognition. *Comput Math Methods Med*. 2014;2014:181245.
47 Sompawong N, Mopan J, Pooprasert P, et al. Automated pap smear cervical cancer screening using deep learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:7044–7048.
48 Hussain E, Mahanta LB, Das CR, et al. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue Cell*. 2020;65:101347.
49 Sanyal P, Ganguli P, Barui S. Performance characteristics of an artificial intelligence based on convolutional neural network for screening conventional Papanicolaou-stained cervical smears. *Med J Armed Forces India*. 2020;76(4):418–424.
50 Holmström O, Linder N, Kaingu H, et al. Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. *JAMA Netw Open*. 2021;4(3):e211740.
51 Ali MM, Ahmed K, Bui FM, et al. Machine learning-based statistical analysis for early stage detection of cervical cancer. *Comput Biol Med*. 2021;139:104985.
52 Lin H, Chen H, Wang X, et al. Dual-path network with synergistic grouping loss and evidence driven risk stratification for whole slide cervical image analysis. *Med Image Anal*. 2021;69:101955.
53 Bhatt AR, Ganatra A, Kotecha K. Cervical cancer detection in pap smear whole slide images using convNet with transfer learning and progressive resizing. *PeerJ Comput Sci*. 2021;7:e348.
54 Gao W, Xu C, Li G, et al. Cervical cell image classification-based knowledge distillation. *Biomimetics (Basel)*. 2022;7(4):195.
55 Kupas D, Harangi B. Classification of pap-smear cell images using deep convolutional neural network accelerated by hand-crafted features. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:1452–1455.
56 Wang W, Tian Y, Xu Y, et al. 3cDe-Net: a cervical cancer cell detection network based on an improved backbone network and multiscale feature fusion. *BMC Med Imaging*. 2022;22(1):130.
57 Alsubai S, Alqahtani A, Sha M, et al. Privacy preserved cervical cancer detection using convolutional neural networks applied to pap smear images. *Comput Math Methods Med*. 2023;2023:9676206.
58 Chowdary GJ, G S, M P, et al. Nucleus segmentation and classification using residual SE-UNet and feature concatenation approach incervical cytopathology cell images. *Technol Cancer Res Treat*. 2023;22:15330338221134833.
59 Kalbhor M, Shinde S, Popescu DE, et al. Hybridization of deep learning pre-trained models with machine learning classifiers and fuzzy min-max neural network for cervical cancer diagnosis. *Diagnostics (Basel)*. 2023;13(7):1363.
60 Tan SL, Selvachandran G, Ding W, et al. Cervical cancer classification from pap smear images using deep convolutional neural network models. *Interdiscip Sci*. 2024;16(1):16–38.
61 Yang T, Hu H, Li X, et al. An efficient Fusion-Purification Network for Cervical pap-smear image classification. *Comput Methods Progr Biomed*. 2024;251:108199.
62 Kyrgiou M, Pouliakis A, Panayiotides JG, et al. Personalised management of women with cervical abnormalities using a clinical decision support scoring system. *Gynecol Oncol*. 2016;141(1):29–35.
63 Lasyk Ł, Barbasz J, Żuk P, et al. An evaluation of the construction of the device along with the software for digital archiving, sending the data, and supporting the diagnosis of cervical cancer. *Contemp Oncol (Pozn)*. 2019;23(3):174–177.
64 Sanyal P, Barui S, Deb P, et al. Performance of a convolutional neural network in screening liquid based cervical cytology smears. *J Cytol*. 2019;36(3):146–151.
65 Tang HP, Cai D, Kong YQ, et al. Cervical cytology screening facilitated by an artificial intelligence microscope: a preliminary study. *Cancer Cytopathol*. 2021;129(9):693–700.
66 Cao L, Yang J, Rong Z, et al. A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. *Med Image Anal*. 2021;73:102197.
67 Li X, Xu Z, Shen X, et al. Detection of cervical cancer cells in whole slide images using deformable and global context aware faster RCNN-FPN. *Curr Oncol*. 2021;28(5):3585–3601.
68 Liang Y, Pan C, Sun W, et al. Global context-aware cervical cell detection with soft scale anchor matching. *Comput Methods Programs Biomed*. 2021;204:106061.
69 Ke J, Shen Y, Lu Y, et al. Quantitative analysis of abnormalities in gynecologic cytopathology with deep learning. *Lab Invest*. 2021;101(4):513–524.
70 Zhu X, Li X, Ong K, et al. Hybrid AI-assistive diagnostic model permits rapid TBS classification of cervical liquid-based thin-layer cell smears. *Nat Commun*. 2021;12(1):3541.
71 Cheng S, Liu S, Yu J, et al. Robust whole slide image analysis for cervical cancer screening using deep learning. *Nat Commun*. 2021;12(1):5639.
72 Kanavati F, Hirose N, Ishii T, et al. A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images. *Cancers (Basel)*. 2022;14(5):1159.
73 Shinde S, Kalbhor M, Wajire P. DeepCyto: a hybrid framework for cervical cancer classification by using deep feature fusion of cytology images. *Math Biosci Eng*. 2022;19(7):6415–6434.
74 Xu C, Li M, Li G, et al. Cervical cell/clumps detection in cytology images using transfer learning. *Diagnostics (Basel)*. 2022;12(10):2477.
75 Alsalatie M, Alquran H, Mustafa WA, et al. Analysis of cytology pap smear images based on ensemble deep learning approach. *Diagnostics (Basel)*. 2022;12(11):2756.
76 Nambu Y, Mariya T, Shinkai S, et al. A screening assistance system for cervical cytology of squamous cell atypia based on a two-step combined CNN algorithm with label smoothing. *Cancer Med*. 2022;11(2):520–529.
77 Du H, Dai W, Zhou Q, et al. AI-assisted system improves the work efficiency of cytologists via excluding cytology-negative slides and accelerating the slide interpretation. *Front Oncol*. 2023;13:1290112.
78 Hamdi M, Senan EM, Awaji B, et al. Analysis of WSI images by hybrid systems with fusion features for early diagnosis of cervical cancer. *Diagnostics (Basel)*. 2023;13(15):2538.
79 Kurita Y, Meguro S, Tsuyama N, et al. Accurate deep learning model using semi-supervised learning and noisy student for cervical cancer screening in low magnification images. *PLoS One*. 2023;18(5):e0285996.
80 Xue P, Xu HM, Tang HP, et al. Improving the accuracy and efficiency of abnormal cervical squamous cell detection with cytologist-in-the-loop artificial intelligence. *Mod Pathol*. 2023;36(8):100186.
81 Chantziantoniou N. BestCyte® primary screening of 500 ThinPrep Pap Test thin-layers: 3 Cytologists' Interobserver diagnostic concordance with predicate manual microscopy relative to Truth Reference diagnoses defining NILM, ASCUS+, LSIL+, and ASCH+ thresholds for specificity, sensitivity, and equivalency grading. *J Pathol Inform*. 2023;14:100182.
82 Arya M, Mittal N, Singh G. Texture-based feature extraction of smear images for the detection of cervical cancer. *IET Comput Vis*. 2018;12(8):1049–1059.
83 Win KP, Kitjaidure Y, Hamamoto K, et al. Computer-assisted screening for cervical cancer using digital image processing of pap smear images. *Appl Sci*. 2020;10(5):1800.

84 Diniz DN, Vitor RF, Bianchi AGC, et al. An ensemble method for nuclei detection of overlapping cervical cells. *Expert Syst Appl*. 2021;185:115642.
85 Sheela Shiney TS, Rose RJ. Deep auto encoder based extreme learning system for automatic segmentation of cervical cells. *IETE J Res*. 2021;69(7):4066–4086.
86 Jia D, Zhou J, Zhang C. Detection of cervical cells based on improved SSD network. *Multimed Tool Appl*. 2021;81(10):13371–13387.
87 Ma D, Liu J, Li J, et al. Cervical cancer detection in cervical smear images using deep pyramid inference with refinement and spatial-aware booster. *IET Image Process*. 2021;14(17):4717–4725.
88 Xiang Y, Sun W, Pan C, et al. A novel automation-assisted cervical cancer reading method based on convolutional neural network. *Biocybern Biomed Eng*. 2020;40(2):611–623.
89 Kim S-H, Oh H-Y, Kim D-W. A study on development of automation diagnosis of liquid based cytology. *Sains Malays*. 2015;44:1729–1738.
90 Hyeon J, Choi H-J, Lee BD, et al. Diagnosing cervical cell images using pre-trained convolutional neural network as feature extractor. In: *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE; 2017.
91 Bai X, Wei J, Starr D, et al. Assessment of efficacy and accuracy of cervical cytology screening with artificial intelligence assistive system. *Mod Pathol*. 2024;37(6):100486.
92 Civit-Masot J, Luna-Perejon F, Muñoz-Saavedra L, et al. A lightweight xAI approach to cervical cancer classification. *Med Biol Eng Comput*. 2024;62(8):2281–2304.
93 Wang J, Yu Y, Tan Y, et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nat Commun*. 2024;15(1):4369.
94 Yang W, Jin X, Huang L, et al. Clinical evaluation of an artificial intelligence-assisted cytological system among screening strategies for a cervical cancer high-risk population. *BMC Cancer*. 2024;24(1):776.
95 Zeng X, Starr D, Li J, et al. AICyte-alone capabilities as an independent screener for triaging cervical cytology using a 50% negative cutoff value. *Cancer Cytopathol*. 2024;132(11):723–730.
96 Mehlhorn G, Kage A, Münzenmayer C, et al. Computer-assisted diagnosis (CAD) in colposcopy: evaluation of a pilot study. *Anticancer Res*. 2012;32(12):5221–5226.
97 Gutiérrez-Fragoso K, Acosta-Mesa HG, Cruz-Ramírez N, et al. Optimization of classification strategies of acetowhite temporal patterns towards improving diagnostic performance of colposcopy. *Comput Math Methods Med*. 2017;2017:5989105.
98 Miyagi Y, Takehara K, Miyake T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images. *Mol Clin Oncol*. 2019;11(6):583–589.
99 Yue Z, Ding S, Li X, et al. Automatic acetowhite lesion segmentation via specular reflection removal and deep attention network. *IEEE J Biomed Health Inform*. 2021;25(9):3529–3540.
100 Cho BJ, Choi YJ, Lee MJ, et al. Classification of cervical neoplasms on colposcopic photography using deep learning. *Sci Rep*. 2020;10(1):13652.
101 Li Y, Chen J, Xue P, et al. Computer-aided cervical cancer diagnosis using time-lapsed colposcopic images. *IEEE Trans Med Imaging*. 2020;39(11):3403–3415.
102 Liu L, Wang Y, Liu X, et al. Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Ann Transl Med*. 2021;9(13):1045.
103 Yu H, Fan Y, Ma H, et al. Segmentation of the cervical lesion region in colposcopic images based on deep learning. *Front Oncol*. 2022;12:952847.
104 Takahashi T, Matsuoka H, Sakurai R, et al. Development of a prognostic prediction support system for cervical intraepithelial neoplasia using artificial intelligence-based diagnosis. *J Gynecol Oncol*. 2022;33(5):e57.
105 Fang S, Yang J, Wang M, et al. An improved image classification method for cervical precancerous lesions based on ShuffleNet. *Comput Intell Neurosci*. 2022;2022:9675628.
106 Kim J, Park CM, Kim SY, et al. Convolutional neural network-based classification of cervical intraepithelial neoplasias using colposcopic image segmentation for acetowhite epithelium. *Sci Rep*. 2022;12(1):17228.
107 Kim S, An H, Cho HW, et al. Pivotal clinical study to evaluate the efficacy and safety of assistive artificial intelligence-based software for cervical cancer diagnosis. *J Clin Med*. 2023;12(12):4024.
108 Chen X, Pu X, Chen Z, et al. Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions. *Cancer Med*. 2023;12(7):8690–8699.
109 Ouh YT, Kim TJ, Ju W, et al. Development and validation of artificial intelligence-based analysis software to support screening system of cervical intraepithelial neoplasia. *Sci Rep*. 2024;14(1):1957.
110 Yue Z, Ding S, Zhao W, et al. Automatic CIN grades prediction of sequential cervigram image using LSTM with multistate CNN features. *IEEE J Biomed Health Inform*. 2020;24(3):844–854.
111 Miyagi Y, Takehara K, Nagayasu Y, et al. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncol Lett*. 2020;19(2):1602–1610.
112 Saini SK, Bansal V, Kaur R, et al. ColpoNet for automated cervical cancer screening using colposcopy images. *Mach Vis Appl*. 2020;31(3):s00138-020-01063-8.
113 Mascarenhas M, Alencoão I, Carinhas MJ, et al. Artificial intelligence and colposcopy: automatic identification of cervical squamous cell carcinoma precursors. *J Clin Med*. 2024;13(10):3003.
114 Aquilina A, Papagiannakis E. Deep learning diagnostic classification of cervical images to augment colposcopic impression. *J Low Genit Tract Dis*. 2024;28(3):224–230.