RGCS REVISTA GREGORIANA DE CIENCIAS DE LA SALUD

ISSN 3028-8940

REVIEW ARTICLE

JULY-DECEMBER 2026;3(2):203-219
https://doi.org/10.36097/rgcs.v3i2.3234

Diagnostic accuracy of artificial intelligence for high-grade cervical lesions in colposcopy: a systematic review

Precisión diagnóstica de la inteligencia artificial en colposcopia para lesiones cervicales de alto grado: una revisión sistemática

Estefanía Bailón-Mieles • Jessica Pilozo

Especialidad en Ginecología y Obstetricia, Escuela de Postgrado, Pontificia Universidad Católica del Ecuador, Sede Manabí, Portoviejo, Manabí, Ecuador.

**Corresponding author*

Recepción: 07-04-2026 *Aceptación:* 15-07-2026 *Publicación:* 31-07-2026

ABSTRACT

Colposcopy has limitations due to observer variability, which is why artificial intelligence (AI) has been proposed to improve its accuracy. This systematic review (PRISMA 2020) analyzed the diagnostic accuracy of AI systems on colposcopic images to detect high-grade cervical intraepithelial lesions (CIN2+, CIN3+, HSIL). Searches were conducted in PubMed, Scopus, Web of Science, ScienceDirect, IEEE Xplore, and Google Scholar (January 2021 – June 2026), including studies with histopathological confirmation as the reference standard and assessing quality with QUADAS-2. Ten studies were included (over 15,000 images and 7,500 patients), mostly retrospective and from Asia. Performance was heterogeneous: sensitivity (63.0–99.9%), specificity (57.6–99.8%), and AUC up to 0.95. AI proved useful as a supportive tool, especially for junior colposcopists. The main biases were related to patient selection and retrospective design. In conclusion, AI shows promising diagnostic accuracy for high-grade lesions, although with significant variability. Prospective, multicenter studies conducted in real-world settings are needed to confirm its effectiveness and cost-effectiveness before widespread clinical implementation.

Keywords: artificial intelligence, colposcopy, cervical intraepithelial neoplasia, high-grade cervical lesions, diagnostic accuracy, cervical cancer.

RESUMEN

La colposcopia tiene limitaciones por la variabilidad del observador, por lo que la inteligencia artificial (IA) se ha propuesto para mejorar su precisión. Esta revisión sistemática (PRISMA 2020) analizó la precisión diagnóstica de sistemas de IA en imágenes colposcópicas para detectar lesiones intraepiteliales cervicales de alto grado (CIN2+, CIN3+, HSIL). Se buscó en PubMed, Scopus, Web of Science, ScienceDirect, IEEE Xplore y Google Scholar (enero de 2021 - junio de 2026), incluyendo estudios con confirmación histopatológica como estándar de referencia y evaluando la calidad con QUADAS-2. Se incluyeron 10 estudios (más de 15.000 imágenes y 7.500 pacientes), mayoritariamente retrospectivos y asiáticos. El rendimiento fue heterogéneo: sensibilidad (63,0-99,9 %), especificidad (57,6-99,8 %) y AUC hasta 0,95. La IA resultó útil como apoyo, especialmente para colposcopistas noveles. Los principales sesgos se relacionaron con la selección de pacientes y el diseño retrospectivo. En conclusión, la IA muestra una precisión diagnóstica prometedora para lesiones de alto grado, aunque con variabilidad significativa. Se necesitan estudios prospectivos, multicéntricos y en condiciones reales para confirmar su efectividad y costo-efectividad antes de su implementación clínica generalizada.

Palabras clave: inteligencia artificial, colposcopia, neoplasia intraepitelial cervical, lesiones de alto grado, precisión diagnóstica, cáncer cervicouterino.

Cite as: Bailón-Mieles, E., & Pilozo, J. (2026). Diagnostic accuracy of artificial intelligence for high-grade cervical lesions in colposcopy: a systematic review. *Revista Gregoriana de Ciencias de la Salud*, *3*(2), 203-219. https://doi.org/10.36097/rgcs.v3i2.3234

INTRODUCTION

Cervical cancer remains a priority public health problem worldwide, despite being a largely preventable neoplasm through vaccination against human papillomavirus (HPV), organized screening, and timely treatment of precursor lesions (World Health Organization [WHO], 2020, 2025). The WHO recognizes it as the fourth most common cancer in women, with approximately 660,000 new cases and 350,000 deaths recorded in 2022, with the greatest disease burden concentrated in low- and middle-income countries (WHO, 2025). According to estimates from the Global Cancer Observatory in 2022, cervical cancer continues to represent a major cause of female morbidity and mortality, especially in regions with less access to prevention, screening, and timely treatment programs (Bray et al., 2024).

The natural history of cervical cancer is closely linked to persistent infection with oncogenic HPV genotypes, particularly types 16 and 18, which are responsible for the majority of invasive cancer cases. Progression from persistent infection to cervical intraepithelial neoplasia (CIN) and subsequently to invasive cancer usually takes several years, providing an opportunity for early detection and treatment of precursor lesions (Perkins et al., 2020). In this context, identifying high-grade cervical intraepithelial lesions, especially CIN2+ and CIN3+, constitutes a fundamental objective of secondary prevention programs, due to their high potential for progression to invasive cancer (WHO, 2020).

Colposcopy is an essential diagnostic tool in the evaluation of women with abnormal screening tests or high-risk HPV infection. Its main function is to identify suspicious epithelial alterations and guide the performance of targeted biopsies for histopathological confirmation (Wentzensen et al., 2015). However, several studies have shown that the diagnostic accuracy of colposcopy depends significantly on the examiner's experience, image quality, and adequate visualization of the transformation zone, leading to considerable inter- and intra-observer variability (Bifulco et al., 2015; Dorji et al., 2022; Vallikad et al., 2017). These limitations can affect the detection of high-grade lesions, particularly in settings with a shortage of trained specialists.

Over the past decade, artificial intelligence (AI) has experienced accelerated growth in the healthcare field, especially in disciplines based on medical image analysis (Rajpurkar et al., 2022). Advances in machine learning and deep learning, particularly through convolutional neural networks, have enabled the development of systems capable of identifying complex patterns with accuracy levels comparable to, or even exceeding, those of human specialists in certain diagnostic areas (Esteva et al., 2017; Topol, 2019). These technologies have demonstrated utility in radiology, dermatology, ophthalmology, and pathology, driving the development of clinical decision support tools (Topol, 2019).

In gynecology, the application of artificial intelligence to the analysis of colposcopic images has emerged as a promising strategy to improve the detection of cervical precursor lesions (Takahashi et al., 2025); recent studies have reported deep learning-based systems capable of classifying colposcopic images, identifying suspicious areas, and differentiating low- and high-grade intraepithelial lesions with high levels of sensitivity and specificity (Wu et al., 2023; Mascarenhas et al., 2024; Ouh et al., 2024). Likewise, it has been observed that the use of AI-based diagnostic assistance systems can improve the performance of less experienced colposcopists and optimize biopsy site selection (Chang et al., 2025).

Despite the growing number of publications, the available evidence shows heterogeneity in terms of the algorithms used, characteristics of the populations studied, methodological quality, outcomes assessed, and diagnostic metrics reported. While some studies analyze the detection of CIN2+, others focus on CIN3+, HSIL, or invasive cancer, making direct comparison of results and determination of the true clinical performance of these technologies difficult (Mascarenhas et al., 2024; Liu et al., 2025; Takahashi et al., 2025).

Therefore, it is necessary to critically synthesize recent scientific evidence related to artificial intelligence applied exclusively to colposcopic images, which is why the present systematic review aimed to analyze the diagnostic accuracy of artificial intelligence systems applied to colposcopic images for the detection of high-grade cervical intraepithelial lesions.

METHODOLOGY

A systematic review of diagnostic accuracy studies was designed, following the recommendations of the PRISMA 2020 statement (Page et al., 2021). The research question was: What is the diagnostic accuracy of artificial intelligence systems applied to colposcopic images for the detection of high-grade cervical intraepithelial lesions?

Studies published between January 2021 and June 2026, in Spanish or English, were included. They had to evaluate artificial intelligence, machine learning, deep learning, convolutional neural network models, or computer-aided diagnostic systems applied to colposcopic images, cervicographic images, or digital images of the cervix. The target population was women undergoing colposcopy due to abnormal cervical screening, high-risk human papillomavirus (HPV) infection, abnormal cytology, or clinical suspicion of cervical lesions.

The primary outcome was diagnostic accuracy for detecting high-grade cervical intraepithelial lesions, defined as CIN2+, CIN3+, HSIL, or cervical cancer. Diagnostic indicators considered were sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy, and area under the curve (AUC). The preferred reference standard was histopathological confirmation via cervical biopsy.

Narrative reviews, systematic reviews, meta-analyses, editorials, letters to the editor, studies without full-text availability, and research focused exclusively on cervical cytology, digital Papanicolaou, computational histopathology, or molecular tests without colposcopic analysis were excluded. Likewise, studies that did not report quantifiable diagnostic metrics or that used only experimental databases without clinical validation were excluded.

The bibliographic search was conducted in PubMed/MEDLINE, Scopus, Web of Science, ScienceDirect, IEEE Xplore, and Google Scholar. Controlled and free terms related to artificial intelligence, machine learning, deep learning, colposcopy, cervicography, cervical intraepithelial neoplasia, high-grade cervical intraepithelial lesions, and diagnostic accuracy were used. The search strategy combined terms such as "artificial intelligence", "machine learning", "deep learning", "convolutional neural network", "colposcopy", "cervical image", "cervicography", "cervical intraepithelial neoplasia", "CIN2+", "CIN3+", and "diagnostic accuracy" using Boolean operators AND and OR. The strategy was adapted to the specific characteristics of each database to maximize the sensitivity and relevance of the retrieved results.

The identified records were exported to Zotero 7 for identification and removal of duplicate records. Subsequently, two reviewers independently evaluated the titles and abstracts of the retrieved articles. Potentially eligible studies were reviewed in full text to determine their final inclusion. Discrepancies between reviewers were resolved through consensus or with the participation of a third reviewer.

Data extraction was performed using a standardized matrix developed in Microsoft Excel. Variables related to author, year of publication, country, journal, study design, sample size, population characteristics, type of image analyzed, type of algorithm used, comparator, reference standard, target lesion, sensitivity, specificity, positive predictive value, negative predictive value, area under the curve, diagnostic accuracy, and main conclusions were collected.

The methodological quality and risk of bias of the included studies were assessed using the QUADAS-2 tool (Whiting et al., 2011), considering four domains: patient selection, index test, reference standard, and flow and timing. Each domain was classified as having a low, high, or unclear risk of bias, according to the criteria established by the authors of the tool.

The included studies were organized and classified according to the type of algorithm used, the diagnostic outcome evaluated, and the reported performance metrics. The results were synthesized through narrative analysis and comparative tables. Due to the expected heterogeneity in the algorithms used, study populations, reference standards, and diagnostic outcomes evaluated, a qualitative synthesis of the evidence was prioritized. However, if sufficient clinical and methodological homogeneity had been identified among the included studies, the possibility of performing a quantitative synthesis of diagnostic indicators was considered.

RESULTS AND DISCUSSION

The bibliographic search performed in PubMed/Medline, Scopus, Web of Science, ScienceDirect, IEEE Xplore, and Google Scholar yielded 148 potentially relevant records. After removing 33 duplicate records, 115 titles and abstracts were screened. Subsequently, 15 articles were selected for full-text assessment. After applying the established eligibility criteria, 5 studies were excluded for not meeting the inclusion criteria. Finally, 10 studies were included in the qualitative synthesis of this review. Figure 1 presents the study selection process.

The 10 included studies were published between 2021 and 2026 and evaluated different applications of artificial intelligence for colposcopic image analysis in the diagnosis of precursor cervical lesions and cervical cancer. Most of the research was conducted in Asia, primarily in China, South Korea, and Japan, while two studies originated in Europe and the United States.

Retrospective diagnostic development and validation designs predominated. However, studies with more robust methodologies were also identified, including a multicenter, double-blind, randomized clinical trial (Kim et al., 2023) and a cross-sectional study of clinical application under real-world conditions (Chang et al., 2025). The sample size varied considerably, ranging from 234 to 7,530 patients, with some studies analyzing thousands of colposcopic images for the training and validation of deep learning algorithms.

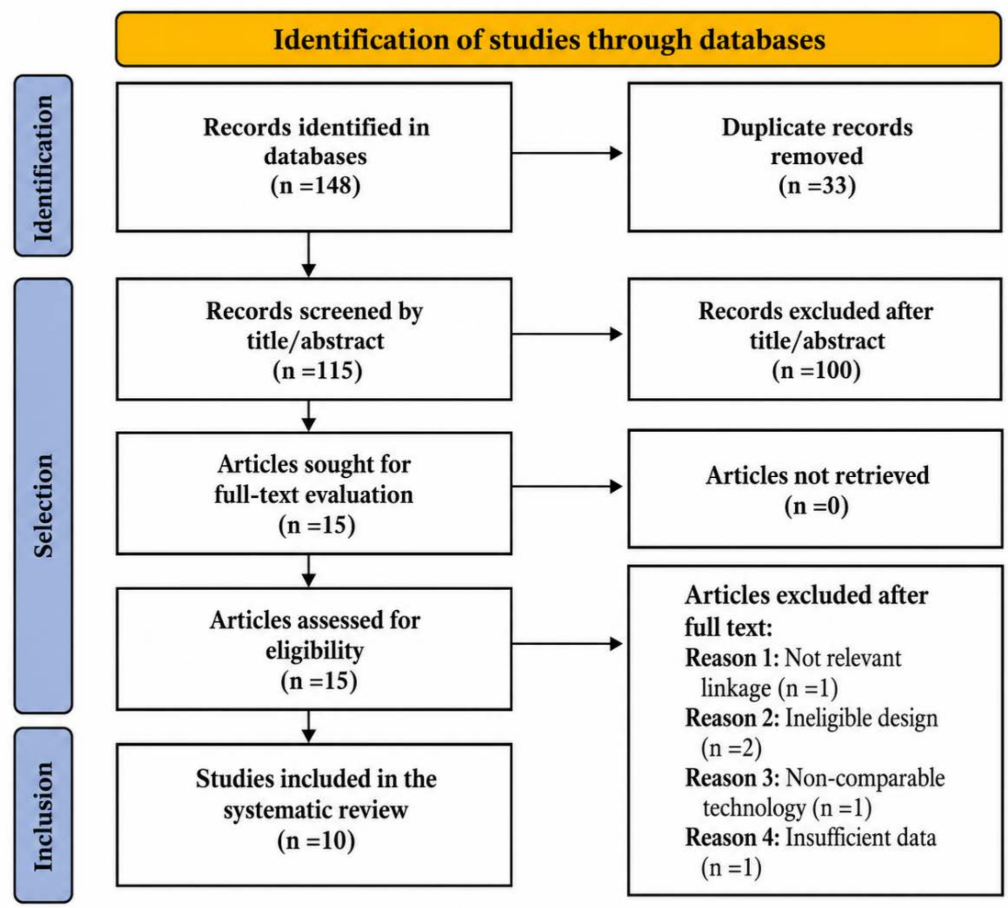

Figure 1. PRISMA 2020 flowchart for the selection of included studies.

The artificial intelligence systems evaluated included convolutional neural networks, deep learning-based models, computer-aided diagnostic systems, and multimodal platforms that integrated clinical variables, cytological results, or human papillomavirus infection status. The outcomes evaluated focused primarily on the detection of high-grade cervical intraepithelial lesions (HSIL, CIN2+, or CIN3+) and cervical cancer. The general characteristics of the included studies are presented in Table 1.

Methodological assessment using the QUADAS-2 tool revealed a low or moderate overall risk of bias in most of the included studies. The domains related to the index test and the reference standard consistently showed a low risk of bias, due to the use of clearly defined algorithms and the use of histopathological confirmation as the diagnostic reference. The main sources of uncertainty were related to patient selection and participant flow, especially in retrospective studies that used pre-selected databases or that did not fully describe losses during the validation process.

Table 1. General and methodological characteristics of the included studies

Reference	Country	Studio design	Sample	AI system/model	Lesion or diagnostic objective
Liu et al. (2021)	China	Retrospective development/ validation	7,530 patients; 15,276 images	CAD based on ResNet50	LSIL+ and HSIL+
Ito et al. (2022)	Japan	Retrospective	463 images	GoogleNet / Inception v1	Normal, CIN1, CIN2-3, invasive cancer
Kim et al. (2022)	South Korea	Retrospective feasibility	234 women	Cerviray AI®	CIN2+ or worse
Zhao et al. (2022)	China	Retrospective diagnostic accuracy	346 women	CAIADS	CIN2+ and CIN3+
Wu et al. (2023)	China	Retrospective external validation	366 women; 1,830 images	CAIADS	CIN2+, CIN3+ and cancer
Kim et al. (2023)	South Korea	Multicenter, double-blind, crossover RCT	886 images out of 7,457	Cerviray AI®	CIN2+/CIN3+
Ouh et al. (2024)	South Korea	Retrospective, multicenter, blinded pivotal trial	400 images	CerviCARE AI; RetinaNet + ResNet-50	P2/P3, HSIL or CIN2+
Aquilina et al. (2024)	Netherlands / United States	Diagnostic development/ validation	4,946 analyzable images	SegFormer + GBT	CIN2+ vs <CIN2
Chang et al. (2025)	China	Cross	825 women	AI-assisted colposcopy	CIN2+ and CIN3+
Booth et al. (2026)	Denmark	Diagnostic development/ validation	178 women; 3,153 images	Cervix -AID-Net	CIN2+ vs <CIN2

Most studies presented a low or uncertain risk of bias in the evaluated domains, with no studies identified as having a high overall risk of bias. The included studies showed considerable variability in the diagnostic performance of artificial intelligence systems applied to colposcopy. In general, the algorithms demonstrated a high capacity for detecting high-grade cervical intraepithelial lesions, with sensitivities generally exceeding 80% and areas under the curve reaching values close to 1.0 in some models.

The best results were observed in systems specifically developed for CIN2+ identification and validated on carefully selected image sets. In contrast, models trained to classify multiple diagnostic categories or evaluated in more heterogeneous clinical scenarios showed more moderate performance. Furthermore, several studies demonstrated that artificial intelligence can function not only as an autonomous diagnostic tool but also as a clinical decision support system, improving the accuracy of less experienced colposcopists. However, direct comparisons between studies should be interpreted with caution due to the heterogeneity of the algorithms used, the characteristics of the populations studied, the reference standards employed, and the diagnostic outcomes evaluated.

Of the 10 included studies, 9 reported comparable diagnostic metrics of sensitivity, specificity, and/or area under the curve, and were therefore incorporated into the comparative synthesis presented in Table 2. Ito et al. (2022) were considered in the qualitative synthesis, but not in the quantitative comparison, because they reported results in multiple diagnostic categories without metrics directly comparable to those used by the other included studies. Detailed results for sensitivity, specificity, and area under the curve for each study are presented in Table 2.

This systematic review demonstrated that artificial intelligence systems applied to the analysis of colposcopic images exhibit favorable diagnostic performance for the detection of high-grade cervical intraepithelial lesions. Most of the included studies reported sensitivities greater than 80%, with areas under the curve reaching values close to 1.0 in some models, suggesting a high discriminatory capacity for identifying precursor lesions with the potential to progress to cervical cancer.

These findings are consistent with recent reviews that have evaluated the role of artificial intelligence in colposcopy. Takahashi et al. (2025), in a systematic review of artificial intelligence applications for the diagnosis of cervical intraepithelial neoplasia and cervical cancer, concluded that deep learning-based models achieve levels of accuracy comparable to those obtained by experienced colposcopists. Similarly, Liu et al. (2025), in a meta-analysis of artificial intelligence applied to the diagnosis of cervical lesions, reported a high overall diagnostic capacity, although they highlighted considerable methodological heterogeneity among the included studies. Similarly, Mascarenhas et al. (2024) noted that artificial intelligence has the potential to reduce the subjectivity inherent in colposcopic interpretation and improve diagnostic reproducibility.

 ISSN 3028-8940 / July-December 2026;3(2):203-219

Table 2. Diagnostic performance of artificial intelligence systems applied to colposcopy for the detection of high-grade cervical intraepithelial lesions

Reference	Target injury	Sensitivity (%)	Specificity (%)	AUC/ Diagnostic accuracy	Reference standard
Liu et al. (2021)	HSIL+	82.3	80.0	AUC: 0.900	Histopathology
Kim et al. (2022)	CIN2+ or worse	74.14	83.05	AUC: 0.777	Histopathology
Zhao et al. (2022)	CIN2+	95.5	57.6	AUC: 0.765	Histopathology
Wu et al. (2023)	CIN2+	80.6	63.5	NR	Histopathology
Kim et al. (2023)	CIN2+/CIN3+	89.18	96.68	AUC IA alone: 0.95; accuracy: 86.40%	Histopathology + expert reference
Ouh et al. (2024)	P2/P3, HSIL or CIN2+	98.0	95.5	NR	Independent committee + histology/cytology
Aquilina et al. (2024)	CIN2+ vs <CIN2	63.0	74.0	AUC: 0.75	Histopathology
Chang et al. (2025)	CIN2+	96.0	NR	NR	Histopathology
Booth et al. (2026)	CIN2+ vs <CIN2	99.9	99.8	Accuracy: 99.8%	Histology of four cervical biopsies

AUC: area under the curve; CIN: cervical intraepithelial neoplasia; HSIL: high-grade squamous intraepithelial lesion; NR: not reported.

The usefulness of artificial intelligence in cervical assessment is not a recent finding. Among the first clinically relevant studies included in this review, Liu et al. (2021) demonstrated that a system based on convolutional neural networks could identify cervical lesions with adequate sensitivity and specificity, while Kim et al. (2022) reported favorable results using an AI-based diagnostic support system for the detection of CIN2+ or higher lesions. These findings are consistent with subsequent research that has shown promising results in the automated analysis of cervical images, such as Hu et al. (2019) developed an automated visual assessment system capable of identifying cervical lesions with performance comparable to that of human experts, while Yuan et al. (2020) demonstrated the feasibility of convolutional neural networks for recognizing cervical squamous lesions in colposcopic images. Similarly, Miyagi et al. (2020) observed that combining colposcopic images with HPV genotype information improved the performance of diagnostic algorithms, suggesting that multimodal models could represent one of the main lines of future development.

A key finding of this review was the considerable variability observed among the included studies. While Booth et al. (2026) reported a sensitivity of 99.9% and a specificity of 99.8%, Aquilina et al. (2024) obtained considerably lower sensitivities. Similarly, Ouh et al. (2024) showed exceptionally high sensitivity and specificity values, exceeding 95%. These differences likely reflect variations in patient selection, image quality, database size, diagnostic criteria used, and the characteristics of the algorithms employed. Several authors have noted that models developed and validated on highly selected image sets tend to perform better than those evaluated in real-world clinical settings (Desai et al., 2022; Liu et al., 2025).

Another relevant aspect identified in this review is that artificial intelligence appears to offer its greatest benefit as a tool to support clinical decision-making rather than as a substitute for specialists. Kim et al. (2023) demonstrated that the incorporation of Cerviray AI® improved diagnostic accuracy during colposcopic evaluation. Similarly, Wu et al. (2023) observed that the CAIADS system improved the performance of less experienced colposcopists, while Chang et al. (2025) documented a significant improvement in sensitivity for the detection of CIN2+ and CIN3+, especially in resource-limited settings. These findings align with the results reported by Ito et al. (2022), who observed an improvement in the diagnostic capabilities of gynecologists when supported by AI-based systems.

From a public health perspective, these results are particularly relevant considering the global goals for the elimination of cervical cancer. The World Health Organization has established strategies aimed at strengthening HPV vaccination, effective screening, and timely treatment of precursor lesions (WHO, 2020, 2021). In this context, artificial intelligence could contribute to optimizing the early detection of high-grade lesions, particularly in regions with limited access to expert colposcopists. Bouvard et al. (2021) highlighted the need to develop innovative technologies that allow for expanding the coverage and quality of cervical screening programs, while Nakisige et al. (2023) noted that the integration of artificial intelligence could represent a viable alternative for improving visual cervical assessment in low- and middle-income countries.

However, the results of this review should be interpreted with caution. Most of the included studies used retrospective designs and pre-selected databases, which may introduce biases related to patient selection and the quality of the analyzed images. Furthermore, there was significant heterogeneity in the outcomes assessed, as some studies used CIN2+ as the primary criterion, while others evaluated CIN3+, HSIL, or invasive cancer. This methodological diversity hindered direct comparison between studies and limited the possibility of conducting a robust quantitative synthesis. For this reason, a narrative synthesis of the available evidence was chosen.

The QUADAS-2 assessment showed that the main risks of bias were related to patient selection and the diagnostic workflow, while the domains corresponding to the index test and the reference standard generally showed a low risk of bias. The use of histopathological confirmation as the reference standard is a significant strength of the included studies, as it allows for an objective evaluation of the diagnostic accuracy of the algorithms.

Taken together, the available evidence suggests that artificial intelligence applied to colposcopy represents a promising tool for improving the detection of high-grade cervical intraepithelial lesions and supporting clinical decision-making. However, before its widespread adoption in clinical practice, prospective, multicenter studies conducted in real-world settings will be necessary to validate its performance, evaluate its cost-effectiveness, and determine its impact on relevant clinical outcomes. Based on the current state of evidence, artificial intelligence should be considered a complementary tool intended to enhance colposcopic interpretation, not a substitute for clinical judgment or histopathological confirmation.

Conclusions

The available evidence indicates that artificial intelligence systems applied to the analysis of colposcopic images demonstrate favorable diagnostic performance for detecting high-grade cervical intraepithelial lesions. Most included studies reported high sensitivity and adequate discriminative ability, suggesting that these technologies can contribute to improving the identification of CIN2+, CIN3+, and HSIL lesions during colposcopic evaluation. Likewise, artificial intelligence showed potential as a clinical decision-support tool, particularly in settings with limited specialist availability or among practitioners with less colposcopy experience. However, the methodological heterogeneity of the studies and the predominance of retrospective designs limit the generalizability of the findings. Consequently, prospective, multicenter research conducted in real-world clinical conditions is required to confirm its clinical utility, assess its cost-effectiveness, and define its role within cervical cancer prevention and early detection programs.

ACKNOWLEDGMENTS

The authors express their gratitude to the Gregorian Journal of Health Sciences for promoting the dissemination of scientific knowledge and creating spaces for publishing research aimed at strengthening the health sciences. They also thank the Pontifical Catholic University of Ecuador for providing the academic environment that fosters the authors' training, research, and professional development.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Estefanía Bailón-Mieles and Jessica Pilozo. **Data curation:** Estefanía Bailón-Mieles and Jessica Pilozo. **Formal analysis:** Estefanía Bailón-Mieles and Jessica Pilozo. **Funding acquisition:** Estefanía Bailón-Mieles and Jessica Pilozo. **Investigation:** Estefanía Bailón-Mieles and Jessica Pilozo. **Methodology:** Estefanía Bailón-Mieles and Jessica Pilozo. **Project administration:** Estefanía Bailón-Mieles and Jessica Pilozo. **Resources:** Estefanía Bailón-Mieles and Jessica Pilozo. **Software:** Estefanía Bailón-Mieles and Jessica Pilozo. **Supervision:** Estefanía Bailón-Mieles and Jessica Pilozo. **Validation:** Estefanía Bailón-Mieles and Jessica Pilozo. **Visualization:** Estefanía Bailón-Mieles and Jessica Pilozo. **Writing – original draft:** Estefanía Bailón-Mieles and Jessica Pilozo. **Writing – review & editing:** Estefanía Bailón-Mieles and Jessica Pilozo.

REFERENCES

Aquilina, A., & Papagiannakis, E. (2024). Deep learning diagnostic classification of cervical images to augment colposcopic impression. *Journal of Lower Genital Tract Disease*, *28*(3), 224. https://doi.org/10.1097/LGT.0000000000000815

Bifulco, G., De Rosa, N., Lavitola, G., Piccoli, R., Bertrando, A., Natella, V., Di Carlo, C., Insabato, L., & Nappi, C. (2015). A prospective randomized study on limits of colposcopy and histology: The skill of colposcopist and colposcopy-guided biopsy in diagnosis of cervical intraepithelial lesions. *Infectious Agents and Cancer*, *10*(1), 47. https://doi.org/10.1186/s13027-015-0042-9

Booth, B. B., Khare, S. K., Blanes-Vidal, V., Bor, P., Nadimi, E. S., & Petersen, L. K. (2026). Higher diagnostic accuracy of an AI model for colposcopy compared with conventional and digital colposcopic evaluation. *Journal of Lower Genital Tract Disease*, *30*(2), 118-123. https://doi.org/10.1097/LGT.0000000000000931

Bouvard, V., Wentzensen, N., Mackie, A., Berkhof, J., Brotherton, J., Giorgi-Rossi, P., Kupets, R., Smith, R., Arrossi, S., Bendahhou, K., Canfell, K., Chirenje, Z. M., Chung, M. H., Pino, M. del, de Sanjosé, S., Elfström, M., Franco, E. L., Hamashima, C., Hamers, F. F., ... Lauby-Secretan, B. (2021). The IARC perspective on cervical cancer screening. *New England Journal of Medicine*, *385*(20), 1908-1918. https://doi.org/10.1056/NEJMsr2030640

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229-263. https://doi.org/10.3322/caac.21834

Chang, Y., Li, T., Zhou, Q., Kang, D., Zhu, L., Yang, J., Kou, Q., He, H., Zhou, Y., Liao, Q., Du, J., Yu, X., & Zhao, Y. (2025). Effectiveness of artificial intelligence-assisted colposcopy in a resource-limited population. *Obstetrics & Gynecology*, 146(4), 545-554. https://doi.org/10.1097/AOG.0000000000006014

Desai, K. T., Befano, B., Xue, Z., Kelly, H., Campos, N. G., Egemen, D., Gage, J. C., Rodriguez, A.-C., Sahasrabuddhe, V., Levitz, D., Pearlman, P., Jeronimo, J., Antani, S., Schiffman, M., & de Sanjosé, S. (2022). The development of automated visual evaluation for cervical cancer screening: The promise and challenges in adapting deep-learning for clinical testing. *International Journal of Cancer*, *150*(5), 741-752. https://doi.org/10.1002/ijc.33879

Dorji, N., Tshering, S., Choden, S., Chhetri, M., Bhujel, D., Wangden, T., Pradhan, B., Bhutia, P. C., & Tshomo, U. (2022). Evaluation of the diagnostic performance of colposcopy in the diagnosis of histologic cervical intraepithelial neoplasia 2+ (CIN2+). *BMC Cancer*, *22*, 930. https://doi.org/10.1186/s12885-022-10030-7

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056

Hu, L., Bell, D., Antani, S., Xue, Z., Yu, K., Horning, M. P., Gachuhi, N., Wilson, B., Jaiswal, M. S., Befano, B., Long, L. R., Herrero, R., Einstein, M. H., Burk, R. D., Demarco, M., Gage, J. C., Rodriguez, A. C., Wentzensen, N., & Schiffman, M. (2019). An observational study of deep learning and automated evaluation of cervical images for cancer screening. *Journal of the National Cancer Institute*, *111*(9), 923-932. https://doi.org/10.1093/jnci/djy225

Ito, Y., Miyoshi, A., Ueda, Y., Tanaka, Y., Nakae, R., Morimoto, A., Shiomi, M., Enomoto, T., Sekine, M., Sasagawa, T., Yoshino, K., Harada, H., Nakamura, T., Murata, T., Hiramatsu, K., Saito, J., Yagi, J., Tanaka, Y., & Kimura, T. (2022). An artificial intelligence-assisted diagnostic system improves the accuracy of image diagnosis of uterine cervical lesions. *Molecular and Clinical Oncology*, *16*(2), 27. https://doi.org/10.3892/mco.2021.2460

Kim, S., An, H., Cho, H.-W., Min, K.-J., Hong, J.-H., Lee, S., Song, J.-Y., Lee, J.-K., & Lee, N.-W. (2023). Pivotal clinical study to evaluate the efficacy and safety of assistive artificial intelligence-based software for cervical cancer diagnosis. *Journal of Clinical Medicine*, *12*(12), 4024. https://doi.org/10.3390/jcm12124024

Kim, S., Lee, H., Lee, S., Song, J.-Y., Lee, J.-K., & Lee, N.-W. (2022). Role of artificial intelligence interpretation of colposcopic images in cervical cancer screening. *Healthcare*, *10*(3), 468. https://doi.org/10.3390/healthcare10030468

Liu, L., Liu, J., Su, Q., Chu, Y., Xia, H., & Xu, R. (2025). Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: A systematic review and meta-analysis. *eClinicalMedicine*, *80*, Article 102992. https://doi.org/10.1016/j.eclinm.2024.102992

Liu, L., Wang, Y., Liu, X., Han, S., Jia, L., Meng, L., Yang, Z., Chen, W., Zhang, Y., & Qiao, X. (2021). Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Annals of Translational Medicine*, *9*(13), 1045. https://doi.org/10.21037/atm-21-885

Mascarenhas, M., Alencoão, I., Carinhas, M. J., Martins, M., Cardoso, P., Mendes, F., Fernandes, J., Ferreira, J., Macedo, G., & Zulmira Macedo, R. (2024). Artificial intelligence and colposcopy: Automatic identification of cervical squamous cell carcinoma precursors. *Journal of Clinical Medicine*, *13*(10), 3003. https://doi.org/10.3390/jcm13103003

Miyagi, Y., Takehara, K., Nagayasu, Y., & Miyake, T. (2020). Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncology Letters*, *19*(2), 1602–1610. https://doi.org/10.3892/ol.2019.11214

Nakisige, C., de Fouw, M., Kabukye, J., Sultanov, M., Nazrui, N., Rahman, A., de Zeeuw, J., Koot, J., Rao, A. P., Prasad, K., Shyamala, G., Siddharta, P., Stekelenburg, J., & Beltman, J. J. (2023). Artificial intelligence and visual inspection in cervical cancer screening. *International Journal of Gynecological Cancer*, *33*(10), 1515-1521. https://doi.org/10.1136/ijgc-2023-004397

Ouh, Y.-T., Kim, T. J., Ju, W., Kim, S. W., Jeon, S., Kim, S.-N., Kim, K. G., & Lee, J.-K. (2024). Development and validation of artificial intelligence-based analysis software to support screening system of cervical intraepithelial neoplasia. *Scientific Reports*, *14*(1), 1957. https://doi.org/10.1038/s41598-024-51880-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(71). https://doi.org/10.1136/bmj.n71

Perkins, R. B., Guido, R. S., Castle, P. E., Chelmow, D., Einstein, M. H., Garcia, F., Huh, W. K., Kim, J. J., Moscicki, A.-B., Nayar, R., Saraiya, M., Sawaya, G. F., Wentzensen, N., Schiffman, M., & 2019 ASCCP Risk-Based Management Consensus Guidelines Committee. (2020). 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors. *Journal of Lower Genital Tract Disease*, *24*(2), 102-131. https://doi.org/10.1097/LGT.0000000000000525

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31-38. https://doi.org/10.1038/s41591-021-01614-0

Takahashi, T., Kobayashi, Y., Sakurai, R., Matsuoka, K., Akatsuka, J., Kisu, I., Iwata, T., Takayama, J., Matsuzaki, M., Yamagami, W., Banno, K., Yamamoto, Y., Matsuoka, H., &

Tamiya, G. (2025). A systematic review of the application of artificial intelligence in colposcopy: Diagnostic accuracy for cervical intraepithelial neoplasia and cervical cancer. *Clinical Medicine Insights: Oncology*, *19*, 11795549251374908. https://doi.org/10.1177/11795549251374908

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Vallikad, E., Siddartha, P. T., Kulkarni, K. A., Firtion, C., Keswarpu, P., Vajinepalli, P., Naik, S., & Gupta, L. (2017). Intra and inter-observer variability of transformation zone assessment in colposcopy: A qualitative and quantitative study. *Journal of Clinical and Diagnostic Research*, *11*(1), XC04-XC06. https://doi.org/10.7860/JCDR/2017/21943.9168

Wentzensen, N., Walker, J. L., Gold, M. A., Smith, K. M., Zuna, R. E., Mathews, C., Dunn, S. T., Zhang, R., Moxley, K., Bishop, E., Tenney, M., Nugent, E., Graubard, B. I., Wacholder, S., & Schiffman, M. (2015). Multiple biopsies and detection of cervical cancer precursors at colposcopy. *Journal of Clinical Oncology*, *33*(1), 83-89. https://doi.org/10.1200/JCO.2014.55.9948

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., Bossuyt, P. M. M., & QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

World Health Organization. (2020). *Global strategy to accelerate the elimination of cervical cancer as a public health problem*. World Health Organization. https://www.who.int/publications/i/item/9789240014107

World Health Organization. (2021). *WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention* (2nd ed.). World Health Organization. https://www.who.int/publications/i/item/9789240030824

World Health Organization. (2025). *Cervical cancer*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/cervical-cancer

Wu, A., Xue, P., Abulizi, G., Tuerxun, D., Rezhake, R., & Qiao, Y. (2023). Artificial intelligence in colposcopic examination: A promising tool to assist junior colposcopists. *Frontiers in Medicine*, *10*, 1060451. https://doi.org/10.3389/fmed.2023.1060451

Yuan, C., Yao, Y., Cheng, B., Cheng, Y., Li, Y., Li, Y., Liu, X., Cheng, X., Xie, X., Wu, J., Wang, X., & Lu, W. (2020). The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific Reports*, *10*(1), 11639. https://doi.org/10.1038/s41598-020-68252-3

Zhao, Y., Li, Y., Xing, L., Lei, H., Chen, D., Tang, C., & Li, X. (2022). The performance of artificial intelligence in cervical colposcopy: A retrospective data analysis. *Journal of Oncology*, *2022*, 4370851. https://doi.org/10.1155/2022/4370851

Statement on the use of artificial intelligence: The authors acknowledge the use of generative artificial intelligence (AI) and AI-assisted technologies solely to improve the readability, language quality, and clarity of this manuscript. AI tools may also have been used to create or edit non-scientific graphical content, such as graphical abstracts, icons, illustrations, diagrams, or educational figures, provided that such use is transparently disclosed and does not misrepresent or fabricate research findings. AI tools were not used to generate, modify, enhance, interpret, or analyze scientific data, medical images (e.g., radiographs, computed tomography scans, magnetic resonance imaging, ultrasound images, histopathological images, microscopy images, or other diagnostic images), or any figures that constitute primary research data. The authors reviewed and edited all AI-assisted content as necessary and assume full responsibility for the accuracy, integrity, and originality of the final published work.