

# Artificial Intelligence in Colposcopy: Diagnostic Performance, Clinical Applications, and Future Directions

Francesco Sesti<sup>a, b</sup>, Carlo Ticconi<sup>a</sup>

## Abstract

**Background:** Colposcopy is a key diagnostic procedure in cervical cancer prevention but is limited by interobserver variability and suboptimal diagnostic accuracy. Artificial intelligence (AI), particularly deep learning–based image analysis, has emerged as a potential adjunct to improve the detection of high-grade cervical lesions. The objective was to review current evidence on the application of AI in colposcopy, with a focus on technical approaches, diagnostic performance, clinical validation, and challenges for implementation in routine practice.

**Methods:** A narrative review informed by PRISMA principles was conducted. PubMed, Embase, and the Cochrane Library were searched for studies evaluating AI-assisted analysis of colposcopic images. Eligible studies reported diagnostic performance outcomes for the detection of cervical intraepithelial neoplasia grade 2 or higher (CIN2+). Study selection is summarized using a PRISMA flow diagram.

**Results:** Most included studies employed convolutional neural network–based models trained on labeled colposcopic image datasets. Reported sensitivities for CIN2+ detection ranged from approximately 62% to over 98%, with specificities between 56% and 98%. Several studies demonstrated diagnostic performance comparable to or exceeding that of experienced colposcopists, particularly in terms of sensitivity. A recent meta-analysis reported pooled sensitivity and specificity of approximately 93% and 85%, respectively.

**Conclusions:** AI-assisted colposcopy shows considerable promise as a tool to enhance diagnostic accuracy and reduce variability in cervical cancer prevention. However, heterogeneity among studies, limited external validation, and challenges related to explainability and clinical integration highlight the need for robust prospective studies before widespread implementation can be recommended.

**Keywords:** Artificial intelligence; Colposcopy; Cervical intraepithelial neoplasia; Cervical cancer; Deep learning; Diagnostic accuracy

## Introduction

Colposcopy plays a central role in cervical cancer prevention, serving as the primary diagnostic procedure for the evaluation of abnormal cervical cytology and high-risk human papillomavirus (HPV) testing results. By enabling visual assessment of the cervix and targeted biopsy of suspicious areas, colposcopy contributes to the detection and management of cervical intraepithelial neoplasia (CIN) and early-stage cervical cancer. However, the diagnostic accuracy of colposcopy is limited by substantial inter- and intra-observer variability, even among experienced practitioners, which can result in missed high-grade lesions or unnecessary biopsies [1, 2].

Reported sensitivities of conventional colposcopy for the detection of CIN2+ vary widely, and specificity remains suboptimal, particularly in low-grade disease [3]. These limitations are amplified in low-resource settings, where access to expert colposcopists is limited.

Artificial intelligence (AI), particularly deep learning approaches, has recently emerged as a potential solution to improve diagnostic consistency and accuracy in image-based medical diagnostics, including colposcopy. Advances in convolutional neural networks (CNNs) have enabled automated analysis of complex visual patterns in medical images, leading to clinically meaningful applications in radiology, dermatology, and ophthalmology [4]. Given the image-driven nature of colposcopy, AI-assisted interpretation of colposcopic images has become an area of growing research interest.

This review aims to summarize and critically appraise current evidence on the application of AI in colposcopy. Specifically, we focus on technical approaches, diagnostic performance for the detection of high-grade cervical lesions, clinical validation, and key challenges that must be addressed before widespread clinical implementation.

## Methods

This narrative review was conducted in accordance with

Manuscript submitted February 3, 2026, accepted February 28, 2026  
Published online March 9, 2026

<sup>a</sup>Section of Gynecology, Department of Surgical Sciences, Tor Vergata University Hospital, Rome, Italy

<sup>b</sup>Corresponding Author: Francesco Sesti, Section of Gynecology, Department of Surgical Sciences, Tor Vergata University Hospital, 00133 Rome, Italy. Email: francesco.sesti@uniroma2.it

doi: <https://doi.org/10.14740/aicm16>  
AI in Clinical Medicine  
(print), 2819-7437 (online)

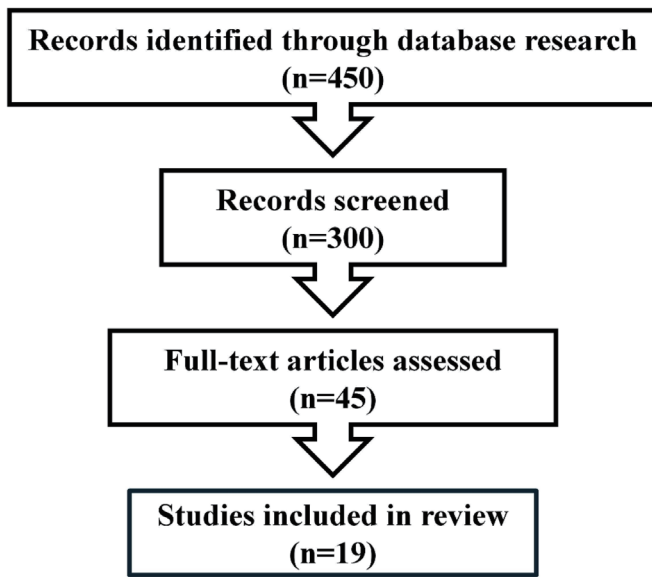


Figure 1. PRISMA flow diagram.

PRISMA principles for transparent reporting of literature identification and selection. A structured search of PubMed, Embase, and the Cochrane Library was performed to identify studies evaluating AI-based applications in colposcopy. The literature identification and study selection process is shown in Figure 1.

The search strategy combined controlled vocabulary terms and free-text keywords related to AI and cervical imaging, including: “artificial intelligence,” “machine learning,” “deep learning,” “convolutional neural network,” “colposcopy,” “cervical intraepithelial neoplasia,” and “cervical cancer.” Searches were limited to studies published in English. Reference lists of included articles and relevant reviews were manually screened to identify additional eligible studies.

Studies were considered eligible if they evaluated AI-based systems applied to colposcopic images and reported diagnostic performance metrics for the detection or classification of cervical lesions, particularly CIN2+ or higher-grade disease. Both retrospective and prospective studies were included. Case reports, editorials, conference abstracts without full data, and studies not reporting diagnostic performance outcomes were excluded.

Data extracted from included studies comprised study design, dataset characteristics, AI methodology, reference

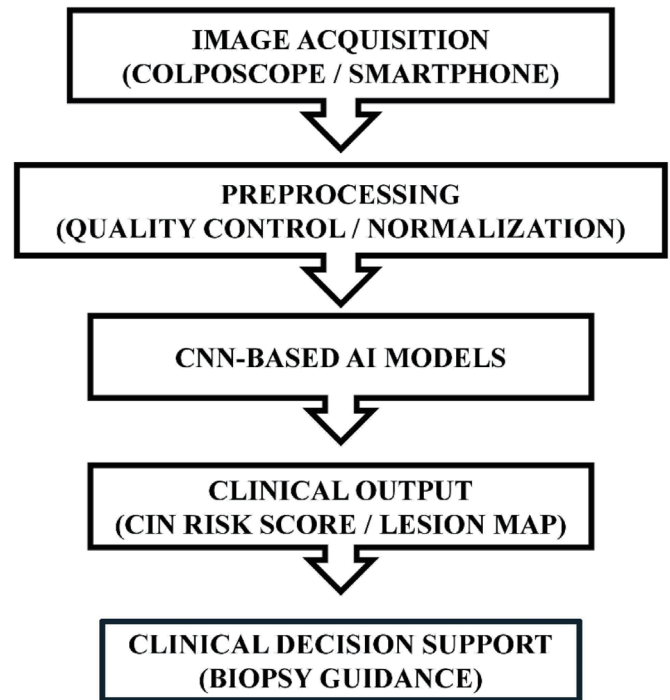


Figure 2. Artificial intelligence (AI)-assisted colposcopy workflow.

standards, performance metrics (including sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC)), and key findings. Due to heterogeneity in study designs, datasets, and outcome measures, a qualitative synthesis was performed rather than a formal meta-analysis. Data extracted from included studies are summarized in Table 1.

## Results

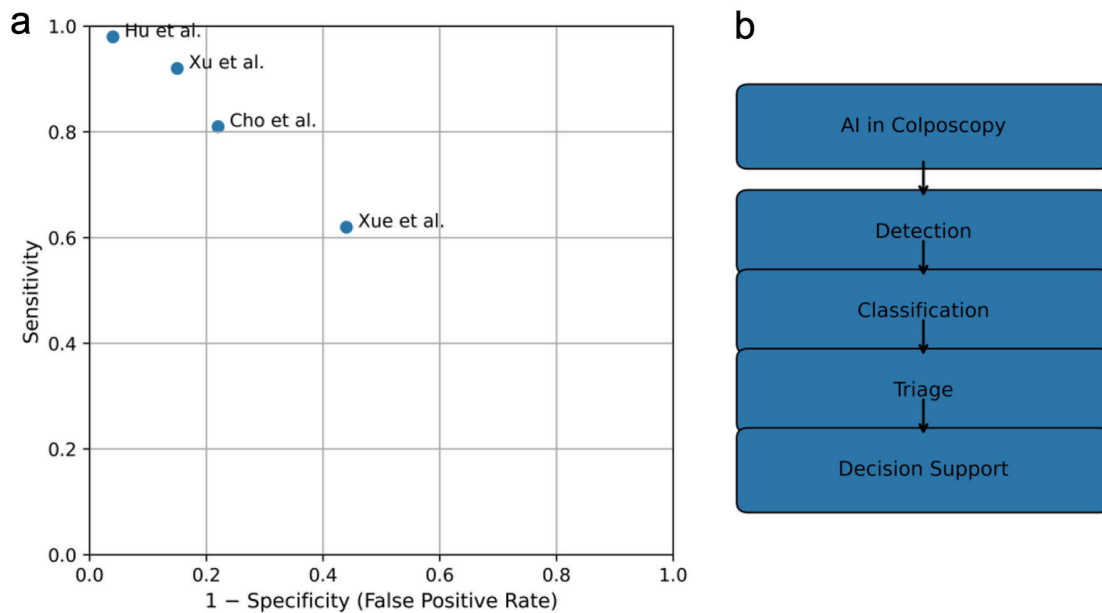
A typical AI-assisted colposcopy pipeline is illustrated in Figure 2. Following image acquisition using standard digital colposcopes or mobile devices, preprocessing steps such as normalization, artifact removal, and color correction are applied to improve image quality and model robustness. The processed images are then analyzed by CNN-based models to generate classification outputs or probability scores for high-grade disease.

The diagnostic performance of AI-assisted colposcopy systems varied across studies, reflecting heterogeneity in data-

Table 1. Summary of Key Studies Evaluating Artificial Intelligence–Assisted Colposcopy

Study	Year	Study design	Dataset	AI method	Target outcome	Performance
Hu et al [5]	2019	Retrospective	9,406 images	CNN	CIN2+ vs. ≤ CIN1	Sensitivity 94.0%, specificity 88.0%, AUC 0.96
Song et al [6]	2020	Retrospective	7,531 images	Deep CNN	CIN2+ detection	Sensitivity 91.3%, specificity 80.0%, AUC 0.93
Xue et al [7]	2020	Retrospective	6,763 images	CNN + attention	CIN2+ detection	Sensitivity 98.0%, specificity 62.0%
Zhao et al [8]	2022	Retrospective	3,673 images	CNN	High-grade CIN	Sensitivity 86.2%, specificity 78.6%
Liu et al [1]	2024	Meta-analysis	33 studies	Multiple models	CIN2+ detection	Pooled sensitivity 93.0%, specificity 85.0%

AUC: area under the receiver operating characteristic curve; CNN: convolutional neural network.



**Figure 3.** Multi-panel synthesis of diagnostic performance and functional roles of artificial intelligence (AI) in colposcopy. (a) Summary receiver operating characteristic (ROC) plot illustrating sensitivity and specificity trade-offs across representative studies evaluating AI-assisted detection of high-grade cervical lesions. Each point represents a study. (b) Flowchart illustrating major clinical task categories of AI in colposcopy, including detection, classification, triage, and clinical decision support.

sets, reference standards, and validation strategies. As summarized in Table 1, reported sensitivities for the detection of CIN2+ ranged from approximately 62% to over 98%, while specificities ranged from 56% to 98% [1, 5–8].

Several comparative studies demonstrated that AI systems achieved diagnostic accuracy comparable to or exceeding that of experienced colposcopists, particularly in terms of sensitivity for high-grade lesions [6, 9]. The diagnostic performance and clinical roles of AI systems are summarized in a multi-panel synthesis (Fig. 3 [5, 7, 9, 10]), illustrating both sensitivity–specificity trade-offs and the range of clinical implementation scenarios.

Notably, many studies relied on internal validation, while fewer conducted external validation on independent datasets. Prospective clinical validation remains limited, underscoring the need for further research before widespread clinical adoption.

## Discussion

### Summary of findings

This review demonstrates that AI-assisted colposcopy has the potential to address key limitations of conventional colposcopic assessment, particularly operator dependence and diagnostic variability. By providing objective and reproducible analysis of colposcopic images, AI systems may support more consistent detection of high-grade cervical lesions.

Most AI systems applied to colposcopy are based on deep learning architectures, particularly CNNs trained on labeled

colposcopic image datasets [5, 6, 8]. These models are designed to recognize visual patterns associated with cervical pathology, such as acetowhite changes, vascular abnormalities, and lesion borders.

Several studies have incorporated explainability techniques, including heatmaps and attention mechanisms, to highlight regions of interest and improve clinician interpretability [9]. In some approaches, clinical variables such as patient age, HPV status, or cytology results are integrated with image-based features to enhance predictive performance.

### Comparison with existing literature

As summarized in Table 1, early large-scale retrospective studies demonstrated high diagnostic performance of CNN-based systems for the detection of high-grade cervical lesions. Hu et al reported a sensitivity of 94.0% and specificity of 88.0% for CIN2+ detection using a CNN trained on more than 9,000 colposcopic images [5]. Similarly, Song et al achieved an AUC of 0.93 with sensitivity exceeding 90% in a dataset of over 7,500 images [6].

Xue et al reported very high sensitivity (98.0%) for CIN2+ detection using a CNN with attention mechanisms, although specificity was comparatively lower (62.0%), highlighting the sensitivity–specificity trade-off observed across studies [7]. Zhao et al demonstrated balanced performance with sensitivity of 86.2% and specificity of 78.6% in a separate retrospective cohort [8].

A recent meta-analysis by Liu et al, including 33 studies, reported pooled sensitivity and specificity of 93.0% and 85.0%, respectively, indicating that AI-assisted colposcopy

can achieve diagnostic performance comparable to experienced colposcopists [1].

### Clinical implications

The findings summarized in Table 1 and Figure 3 suggest that AI systems can achieve high sensitivity for CIN2+ detection, which is critical for cervical cancer prevention. Such systems may be particularly valuable in low-resource settings, where access to expert colposcopists is limited and diagnostic support tools could have a significant clinical impact.

Despite promising results, several challenges remain. Many studies included in this review were retrospective and used curated datasets that may not reflect real-world clinical variability. Furthermore, issues related to model generalizability, explainability, regulatory approval, and integration into clinical workflows must be addressed before routine clinical implementation can be recommended [11].

### Barriers to clinical translation of AI in colposcopy

Despite the promising diagnostic performance reported across multiple studies, several important barriers continue to limit the clinical translation of AI-assisted colposcopy into routine practice.

#### *Limited dataset diversity and risk of overfitting*

One of the most significant challenges is the limited diversity of datasets used for model training and validation. Many studies rely on retrospective image collections obtained from single institutions or geographically restricted populations. Such datasets may not adequately represent the variability encountered in real-world clinical settings, including differences in patient demographics, HPV genotype distribution, cervical anatomy, image acquisition protocols, and colposcope hardware. As a result, AI models may learn features specific to the training environment rather than generalizable disease-related patterns, increasing the risk of overfitting.

Additionally, inadequate separation between training and testing datasets can introduce data leakage, particularly when images from the same patient or examination are inadvertently included in both sets. This can artificially inflate performance metrics and overestimate clinical utility. External validation using independent, prospectively collected datasets remains limited but is essential to establish the robustness and generalizability of AI systems.

#### *Limitations of histopathology as the reference standard*

Histopathological assessment of cervical biopsy specimens is widely used as the reference standard for evaluating diagnostic performance; however, it is not a perfect gold standard. Biopsy sampling may miss the most severe lesion due to incomplete

sampling, lesion heterogeneity, or suboptimal targeting during colposcopy. Interobserver variability among pathologists further contributes to diagnostic uncertainty, particularly in distinguishing between low-grade and high-grade lesions.

These limitations introduce inherent noise into training labels, which may affect AI model performance and reliability. Furthermore, AI systems trained on imperfect reference standards may replicate existing diagnostic biases rather than overcome them. This highlights the importance of incorporating standardized pathology review and, where feasible, consensus diagnosis or longitudinal clinical outcomes in future validation studies.

#### *Uncertainty in clinical interpretation of AI probability scores*

Another critical barrier is the lack of clarity regarding how AI-generated probability scores should be interpreted and integrated into clinical decision-making. While many studies report model outputs in terms of predicted probabilities for CIN2+ or high-grade disease, clear thresholds for clinical action have not been standardized. Clinicians must determine how these probabilities translate into practical decisions, such as whether to perform a biopsy, increase surveillance, or defer intervention.

Without clearly defined decision thresholds and clinical integration pathways, AI outputs may remain difficult to interpret and apply in routine practice. Prospective studies evaluating AI-assisted decision-making in real-world clinical workflows are needed to determine whether AI improves patient outcomes, reduces unnecessary biopsies, or enhances diagnostic consistency.

#### *Lack of prospective clinical validation and workflow integration*

Most existing studies are retrospective and focus primarily on algorithm development and internal validation. Prospective clinical trials assessing the impact of AI-assisted colposcopy on diagnostic accuracy, workflow efficiency, and patient outcomes remain limited. Furthermore, integration into clinical workflows requires consideration of usability, interpretability, regulatory approval, and clinician acceptance.

Successful clinical translation will depend not only on diagnostic performance but also on the ability of AI systems to function as reliable, transparent, and user-friendly decision-support tools within routine colposcopic practice.

### Explainable AI and alignment with colposcopic diagnostic criteria

A critical requirement for the clinical adoption of AI in colposcopy is the interpretability of model predictions. Explainable AI (XAI) methods provide visual and quantitative tools that allow clinicians to understand the basis of algorithmic decision-making, thereby facilitating validation, regulatory approval,

and clinical trust. In colposcopic image analysis, explainability is most commonly implemented through saliency maps, Gradient-weighted Class Activation Mapping (Grad-CAM), attention maps, and feature attribution techniques.

These visualization methods generate heatmaps highlighting image regions that most strongly influence the model's prediction. For clinical relevance, it is essential that these highlighted regions correspond to established diagnostic features used in standard colposcopic assessment, such as acetowhite epithelium, mosaic patterns, punctuation, lesion margins, and vascular abnormalities. These features form the basis of widely accepted scoring systems such as Reid's Colposcopic Index, which integrates lesion color, margins, vascular patterns, and iodine staining characteristics to estimate lesion severity.

Several recent studies that incorporate explainable AI methods have demonstrated promising alignment between AI-generated activation maps and clinically relevant lesion regions in colposcopic image analysis. For example, explainability techniques including Grad-CAM and saliency maps have been used to visualize the regions that drive model predictions, showing that AI attention corresponds closely with annotated lesion areas and clinically salient features such as acetowhite epithelium and abnormal vascular patterns [12, 13].

Explainable AI also serves an important role in identifying potential model failure modes. Visualization techniques can reveal when models focus on irrelevant regions, such as image borders, specular reflections, or background artifacts, which may indicate overfitting or dataset bias. Such insights enable model refinement and improve reliability.

Furthermore, explainability facilitates integration into clinical workflows by enabling clinicians to verify AI recommendations in real time. Rather than functioning as opaque "black-box" systems, interpretable AI tools can support clinical reasoning by providing transparent visual justification aligned with established diagnostic criteria. This human-AI interpretability alignment is essential for clinician acceptance, regulatory approval, and safe clinical implementation.

Future research should prioritize standardized evaluation frameworks to quantify the agreement between AI-generated heatmaps and expert-defined lesion regions. Prospective studies incorporating clinician feedback and explainability-guided model validation will be essential to ensure that AI systems support, rather than obscure, clinical decision-making.

### Future directions

Despite substantial advances in AI-assisted colposcopy, several critical gaps remain that must be addressed to enable safe and effective clinical implementation. We propose the following priorities for future research:

1) Prospective, multicenter trials. To assess generalizability across diverse populations, image acquisition protocols, and clinical settings. These studies will provide robust evidence on real-world performance and integration feasibility.

2) Standardized benchmarking datasets. The development of open-access, well-annotated colposcopic image repositories with diverse patient demographics and clear reference standards will facilitate transparent comparison of AI models and

reproducibility of findings.

3) Intervention and outcome studies. Evaluating the impact of AI-assisted colposcopy on clinical decision-making, biopsy rates, detection of high-grade lesions, patient outcomes, and cost-effectiveness is essential to move beyond algorithmic performance metrics.

4) Human-AI collaboration models. Investigating how AI outputs can be optimally integrated with clinician judgment, including explainable AI methods, decision-support tools, and workflow integration, will maximize adoption and trust.

5) Regulatory and ethical considerations. Research into regulatory pathways, patient consent, data privacy, and bias mitigation is necessary to ensure responsible deployment of AI in cervical cancer screening.

By addressing these priorities, the field can transition from proof-of-concept studies to clinically actionable AI solutions that improve cervical cancer detection and patient care.

## Conclusions

AI represents a promising adjunct to colposcopy, with the potential to improve diagnostic performance, reduce variability, and support clinical decision-making in cervical cancer prevention. Current evidence suggests that AI-assisted systems can achieve high sensitivity and competitive accuracy for the detection of high-grade cervical lesions. Nevertheless, robust external validation, prospective clinical evaluation, and thoughtful integration into clinical practice are essential before widespread implementation can be recommended.

## Acknowledgments

None to declare.

## Financial Disclosure

The authors declare no funding or financial support for this work.

## Conflict of Interest

The authors declare no conflict of interest.

## Informed Consent

Not applicable.

## Author Contributions

Francesco Sesti: conceptualization and study design, writing – original draft and manuscript revision, literature search, and data extraction. Carlo Ticconi: analysis and synthesis of AI and

colposcopy studies, preparation of figures and tables, and approval of the final version for submission.

## Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

## References

1. Liu L, Liu J, Su Q, Chu Y, Xia H, Xu R. Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: a systematic review and meta-analysis. *EClinicalMedicine*. 2025;80:102992. [doi pubmed](#)
2. Massad LS, Einstein MH, Huh WK, Katki HA, Kinney WK, Schiffman M, Solomon D, et al. 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. *J Low Genit Tract Dis*. 2013;17(5 Suppl 1):S1-S27. [doi pubmed](#)
3. Oliveira CA, et al. Accuracy of colposcopy in cervical cancer screening: a systematic review. *Int J Gynecol Cancer*. 2018;28:1-7.
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. [doi pubmed](#)
5. Hu L, Bell D, Antani S, Xue Z, Yu K, Horning MP, Gachuhi N, 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. [doi pubmed](#)
6. Song B, et al. Deep learning-based classification of cervical lesions using colposcopic images. *Gynecol Oncol*. 2020;158:668-675.
7. Xue P, Ng MT, Qiao Y, et al. Deep learning-based detection of cervical precancerous lesions using attention mechanisms. *Med Image Anal*. 2020;65:101732.
8. Zhao Y, et al. Performance of artificial intelligence in cervical colposcopy: a retrospective study. *Gynecol Oncol*. 2022;165:640-648.
9. Xu T, Zhang H, Xin Y, et al. Deep learning-based assessment of cervical lesions using colposcopic images: a multicenter study. *Int J Cancer*. 2021;149:1819-1829.
10. Cho BJ, Baek JH, Lee HS, et al. Development and validation of a deep learning-based artificial intelligence system for colposcopic diagnosis of cervical neoplasia. *Scientific Reports*. 2021;11:13608. [doi](#)
11. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. [doi pubmed](#)
12. Novitasari D, Nugroho H, Nugraha Y, et al. Explainable artificial intelligence for colposcopic image analysis: Interpretable deep learning models for cervical lesions. *Expert Systems with Applications*. 2024;216:119820.
13. Li J, Wang Y, Zhou X, et al. Deep learning-based segmentation and heatmap visualization of cervical lesion regions using colposcopic images. *Heliyon*. 2023;9(3):e13892.