INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

REVIEW ARTICLE

Application of Artificial Intelligence in HPV Related Cervical Cancer Diagnosis: A Systematic Review

Tumpa Dey Barua[1]*

1 MS (Obstetrics & Gynaecology), Chattogram Medical College Hospital, Chattogram, Bangladesh

OPEN ACCESS

Received: 17/06/2026

Accepted: 26/07/2026

Published: 17/08/2026

Citation: Barua TD (2026) Application of Artificial Intelligence in HPV Related Cervical Cancer Diagnosis: A Systematic Review. Indian Journal of Science and Technology 19(30): 2119-2130. https://doi.org/10.17485/IJST/v19i30.919

* **Corresponding author**.

drtumpadeybarua@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Barua. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: This review aims to quantify AI diagnostic performance specifically for HPV-related cervical cancer detection, stratified by methodological category, and to evaluate the added value of HPV genotype integration. **Methods:** A systematic review was conducted following PRISMA 2020 guidelines. SciSpace, PubMed/MEDLINE, Google Scholar, and ArXiv were searched for records published between January 2015 and May 2025, yielding 647 records; after deduplication (n=450) and two-stage screening, 37 studies met inclusion criteria. Two reviewers independently extracted study design, AI methodology, dataset characteristics, and performance metrics; quality was appraised using QUADAS-2 and PROBAST. **Findings:** Deep learning dominated the included literature (24/37, 64.9%), followed by classical machine learning (8/37, 21.6%) and multimodal fusion (5/37, 13.5%). Reported accuracy spanned 83.0-99.99%, sensitivity 80.0-100%, specificity 67.0-99.0%, and AUC 0.85-0.96. Multimodal models fusing image data with HPV genotype information achieved the highest ceiling performance (AUC up to 0.963), exceeding image-only deep learning and expert-clinician benchmarks reported within the same studies. **Novelty:** Unlike prior broad reviews of AI in cervical cancer, this is the first systematic review to isolate HPV genotype-linked diagnostic AI as a distinct analytic unit, the first in this space to apply dual formal risk-of-bias appraisal (QUADAS-2, PROBAST) across all included studies, and the first to quantify the performance differential between genotype-fused multimodal models and image-only approaches (AUC up to 0.963 versus 0.96). This stratification provides a quantitative framework for evaluating the added value of multimodal architectures, synthesised into a translational roadmap for future AI development.

Keywords: Artificial Intelligence; Cervical Cancer; Human Papillomavirus; Deep Learning; Machine Learning; Convolutional Neural Network; Systematic Review

1 Introduction

Cervical cancer continues to rank as the fourth most frequently diagnosed cancer and the fourth leading cause of cancer death among women globally, with an estimated 604,000 incident cases and 342,000 deaths recorded in 2020 [1]. The burden is markedly skewed toward low- and middle-income countries, where organised screening infrastructure is often unavailable; in South Asian settings, cervical cancer remains among the leading causes of cancer mortality in women [2]. Persistent infection with high-risk HPV genotypes, principally HPV16 and HPV18, is now recognised as a necessary precursor to nearly all cervical cancers, accounting for roughly 70% of cases [2]. Disease progression follows a well-characterised sequence through cervical intraepithelial neoplasia (CIN) grades 1 through 3 before invasive carcinoma develops, and intervention at the precancerous stage offers the greatest chance of cure [3]. Genomic profiling work by Long et al. [4] identified gene panels associated with cervical carcinogenesis that continue to inform the design of data-driven early-diagnosis models.

Cytology-based Pap smear testing and HPV DNA testing remain the mainstay of conventional screening, but both carry recognised limitations. Cytology is labour-intensive, depends on trained cytopathologists, and is subject to inter-observer sensitivity reported as low as 51% and as high as 85% [5]. HPV DNA testing offers high sensitivity but limited specificity, requiring supplementary triage to avoid unnecessary colposcopy referral [6]. In resource-constrained settings, workforce and laboratory shortages compound these problems [7]. Early efforts to formalise computational triage, such as the hierarchical modular neural network architecture proposed by Ali et al. [8], demonstrated more than a decade ago that automated risk stratification was technically feasible, foreshadowing the current wave of AI-based diagnostic research.

Artificial intelligence, spanning classical machine learning and deep learning, has since matured into a credible complement to conventional screening. Convolutional neural networks (CNNs) now match or exceed expert clinicians on selected image-classification benchmarks in pathology and radiology [9], [10], and in cervical cancer specifically, AI has been applied across cytology image analysis, colposcopy classification, HPV genotype-based risk prediction, and multimodal data fusion [11], [12], [13] (Figure 1).

Several reviews of AI in cervical cancer already exist, but each leaves the HPV genotype-specific question unresolved. Tulimilli et al. [14] and Silva et al. [15] each provide broad narrative surveys spanning screening, diagnosis, prognosis, and treatment, but neither restricts scope to HPV-linked diagnostic pipelines nor reports pooled, methodology-stratified accuracy ranges. Kondo et al. [16] systematically reviewed AI applications built on cervical cancer risk-factor data rather than imaging, leaving image-based and multimodal diagnostic evidence largely unexamined. Giansanti et al. [17] offers a narrative account of AI in cervical cytology specifically, without a PRISMA-defined search or formal risk-of-bias appraisal. Priya et al. [18] conducted a meta-analysis restricted to colposcopy-based models, excluding cytology and multimodal genotype-integrated approaches, while Takahashi et al. [19] similarly narrowed their systematic review to colposcopy diagnostic accuracy alone. Dellino et al. [20] and William and Ware [21] discuss opportunities and implementation barriers at a policy level rather than synthesising quantitative diagnostic performance. Pawar et al. [13] provide a broad overview of AI-driven screening and treatment strategies without a systematic search methodology, and Wang et al. [22] survey colposcopic and cytological image analysis broadly without isolating genotype-fusion performance or applying formal bias appraisal. Nguyen et al. [23] further characterised deep learning applications in cervical cancer cell diagnosis and screening, again without a genotype-specific or quality-appraised synthesis. Across these ten reviews, the specific question of how much diagnostic value HPV genotype integration adds relative to image-only AI models remains unquantified, and dual formal quality appraisal is applied in none of them.

Given these gaps, we conducted the present systematic review. Sections 1.1 to 1.3 state, respectively, the contributions of this review, the specific literature gaps it addresses, and the research questions guiding data extraction and synthesis.

1.1 Research Contributions

This review makes four contributions relative to the existing literature. First, it applies a PRISMA 2020-compliant search across four databases restricted specifically to HPV-linked cervical cancer diagnostic AI, rather than to cervical cancer AI broadly. Second, it applies dual formal quality appraisal, QUADAS-2 for diagnostic accuracy studies and PROBAST for prediction models, to all 37 included studies, a step largely absent from prior narrative reviews in this space [14], [15], [17]. Third, it reports methodology-stratified performance bands across deep learning, classical machine learning, and multimodal approaches rather than aggregate figures, enabling direct comparison of image-only versus genotype-fused models. Fourth, it consolidates the synthesis into milestone, gap, and bottleneck summary tables intended as a practical reference for researchers designing the next generation of HPV-integrated diagnostic models.

1.2 Research Gaps

Three gaps motivate this review. First, existing reviews of AI in cervical cancer [2], [3], [4], [5], [10] address the disease broadly rather than isolating HPV genotype-linked diagnostic evidence, obscuring the specific contribution of genotype data to model performance. Second, formal risk-of-bias appraisal is inconsistently applied; several reviews [17], [20], [21] are narrative in structure and report neither a PRISMA flow diagram nor a quality-assessment tool, limiting confidence in the strength of the underlying evidence base. Third, no prior review has consolidated methodology-specific accuracy, sensitivity, specificity, and AUC ranges into a single comparative summary capable of informing model-selection decisions for future HPV-related diagnostic tool development.

1.3 Research Questions

This review addresses three questions: (1) What methodological approaches, namely deep learning, classical machine learning, and multimodal integration, have been applied to HPV-related cervical cancer diagnosis, and how are they distributed across the published literature? (2) What diagnostic performance, measured by accuracy, sensitivity, specificity, and AUC, is associated with each methodological category? (3) Does the integration of HPV genotype data with image-based AI models improve diagnostic performance relative to image-only approaches, and what translational barriers remain before clinical deployment?

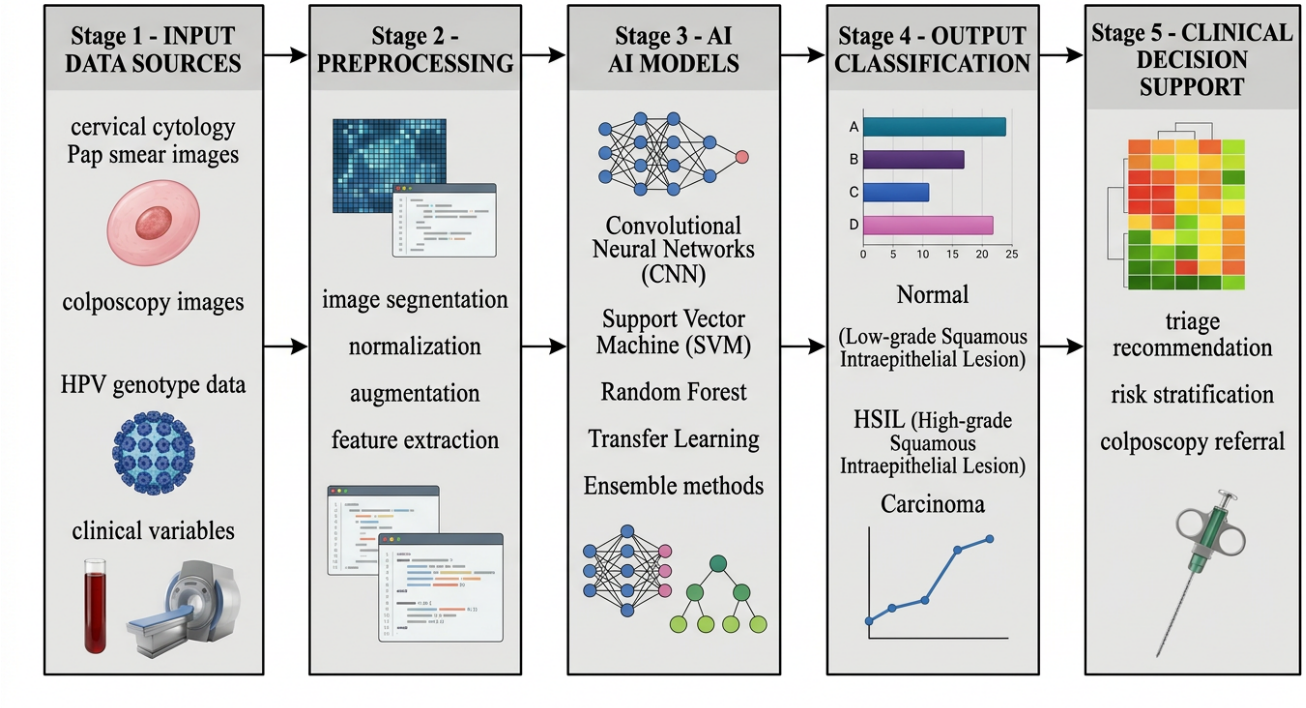


Fig 1. AI pipeline workflow for HPV related cervical cancer diagnosis, illustrating the five sequential stages from data input through clinical decision support.

2 Review Methodology

2.1 Search Strategy

This systematic review was conducted in accordance with PRISMA 2020 reporting guidelines. Four electronic databases were searched: SciSpace, PubMed/MEDLINE, Google Scholar, and ArXiv. The search, conducted in May 2025, combined the terms ”artificial intelligence”, ”machine learning”, ”deep learning”, ”convolutional neural network”, ”HPV”, ”human papillomavirus”,

”cervical cancer”, ”cervical intraepithelial neoplasia”, ”Pap smear”, ”colposcopy”, ”cytology”, ”diagnosis”, ”screening”, and ”detection”, combined using Boolean operators. No language restriction was applied.

The four databases were selected to balance comprehensive biomedical coverage with recall of AI-specific and grey literature relevant to this interdisciplinary topic. PubMed/MEDLINE indexes the large majority of peer-reviewed clinical and biomedical journals relevant to this review and provided the core evidence base for the clinical domain. Google Scholar was searched specifically to capture conference proceedings and AI-venue publications that are a common publication route for applied deep learning work but are inconsistently indexed in subscription bibliographic databases. ArXiv was searched to capture AI/ML preprints, reflecting the short preprint-to-publication lag typical of this field. SciSpace was used as an AI-assisted semantic search layer to cross-check Boolean retrieval across the other three sources. Scopus and Web of Science were not searched directly; the resulting indexing-coverage risk is addressed as a limitation in Section 4.

To assess whether the evidence base had materially changed since the original search, a supplementary, non-systematic check of PubMed/MEDLINE and Google Scholar was performed in July 2026 using the same search terms. This check surfaced additional narrative and systematic reviews of AI in cervical cancer published in early-to-mid 2026 covering overlapping ground, but did not identify additional primary studies reporting HPV genotype-stratified diagnostic performance that would meet the eligibility criteria in Section 2.2. The methodology-stratified performance ranges reported in Section 3 are therefore considered current as of the original search date; this is noted as a limitation in Section 4.

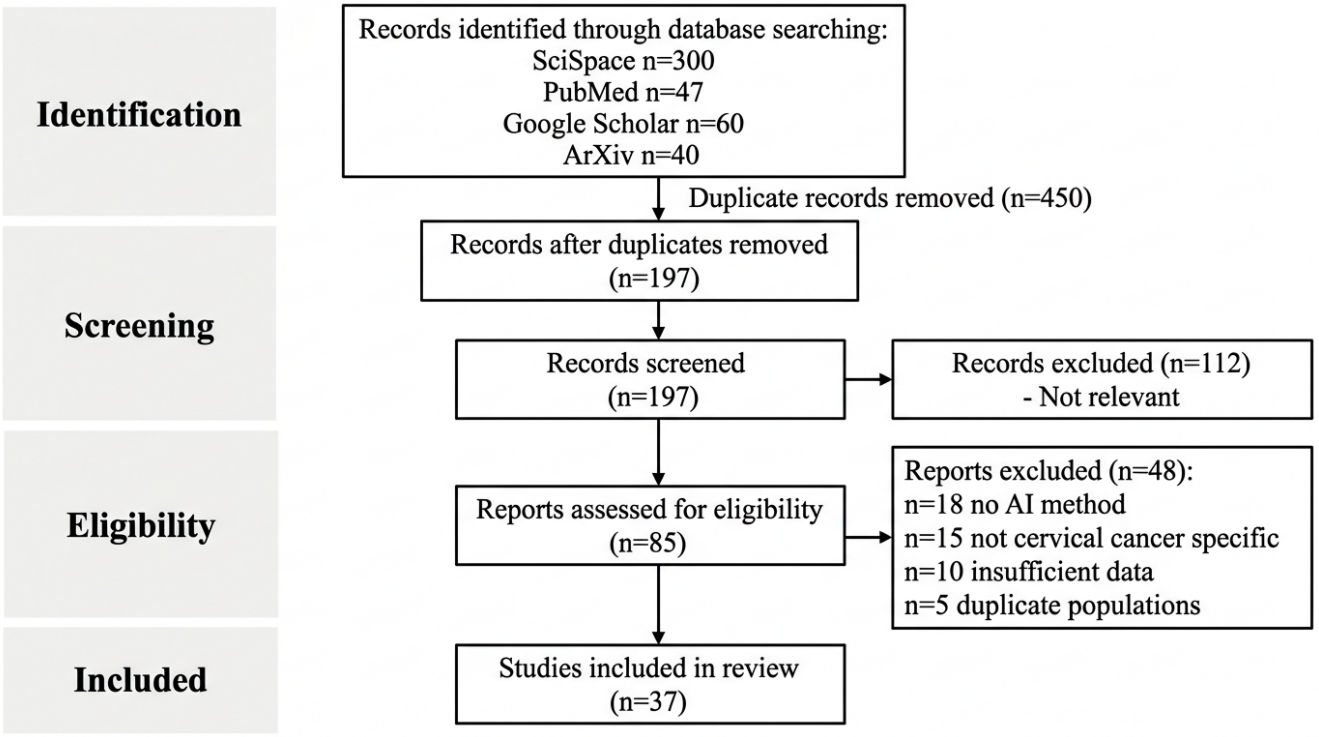


Fig 2. PRISMA 2020 flow diagram illustrating the systematic search and study selection process. A total of 647 records were identified across SciSpace, PubMed, Google Scholar, and ArXiv. After removal of 450 duplicates, 197 unique records were screened by title and abstract. Of these, 112 were excluded, and 85 full-text articles were assessed for eligibility. Following full-text review, 48 articles were excluded (no AI method, n=18; not specific to cervical cancer or HPV, n=15; insufficient data, n=10; duplicate population, n=5). Thirty-seven studies were included in the final qualitative synthesis.

2.2 Inclusion and Exclusion Criteria

Studies were included if they: (1) reported original research on AI applications in cervical cancer or HPV-related diagnosis; (2) employed machine learning, deep learning, or related computational methods; (3) reported quantitative performance metrics; (4) were published between January 2015 and May 2025; and (5) were available in full text. Studies were excluded if they did

not employ an AI method (n=18), were not specific to cervical cancer or HPV (n=15), provided insufficient data for quality assessment (n=10), or represented duplicate populations drawn from the same cohort (n=5).

2.3 Study Selection, Data Extraction, and Quality Assessment

Two independent reviewers screened titles and abstracts of the 197 unique records retrieved after deduplication, with full-text assessment performed for the 85 studies that passed initial screening. Data were extracted using a standardised form capturing study design, AI methodology, dataset characteristics, sample size, performance metrics (accuracy, sensitivity, specificity, AUC), and key findings. Methodological quality was appraised using the QUADAS-2 tool for diagnostic accuracy studies and the PROBAST checklist for prediction-model studies. References [4], [24], and [25] are cited in this review as contextual and background sources; they were not among the 37 PRISMA-included studies shown in Figure 2.

3 Overview

This section synthesises the evidence extracted from the 37 included studies. Section 3.1 reports study characteristics and methodology-stratified findings; Section 3.2 highlights milestone studies and technological breakthroughs; and Section 3.3 consolidates the research gaps, bottlenecks, and emerging directions identified across the reviewed literature.

3.1 Results and Findings

3.1.1. Study Characteristics

The 37 included studies were published between 2017 and 2025, with the majority (n=29, 78.4%) published from 2020 onward, reflecting the rapid acceleration of AI research in this field. Studies originated from diverse geographic settings, including China (n=12, 32.4%), the United States (n=8, 21.6%), Europe (n=7, 18.9%), South and Southeast Asia (n=5, 13.5%), and other regions (n=5, 13.5%) (Figure 3). Study designs included retrospective cohort studies (n=18, 48.6%), cross-sectional diagnostic studies (n=11, 29.7%), prospective studies (n=4, 10.8%), systematic reviews (n=3, 8.1%), and protocol papers (n=1, 2.7%) (Figure 4). The predominance of retrospective designs reflects the early stage of clinical translation in this field, where algorithm development and internal validation have outpaced prospective and real-world evaluation. Sample sizes ranged from 150 to 21,720 participants, with image dataset sizes ranging from 424 to 17,013 images. Full study-level characteristics for all 37 included studies are provided in Supplementary Table S1; principal datasets are summarised in Table 2.

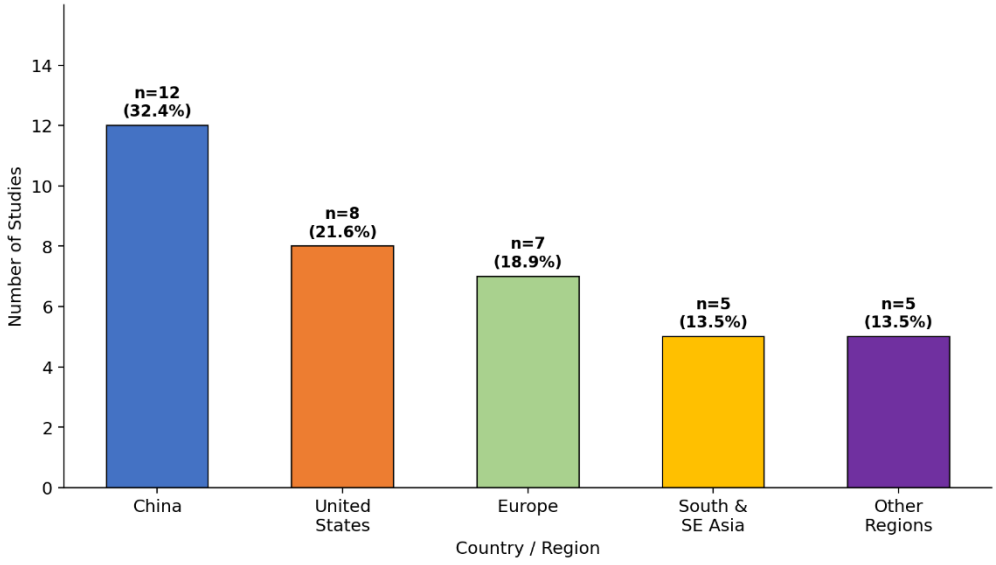


Fig 3. Geographic distribution of included studies by country of origin. China contributed the largest proportion of studies (n=12, 32.4%), followed by the United States (n=8, 21.6%), Europe (n=7, 18.9%), South and Southeast Asia (n=5, 13.5%), and other regions (n=5, 13.5%).

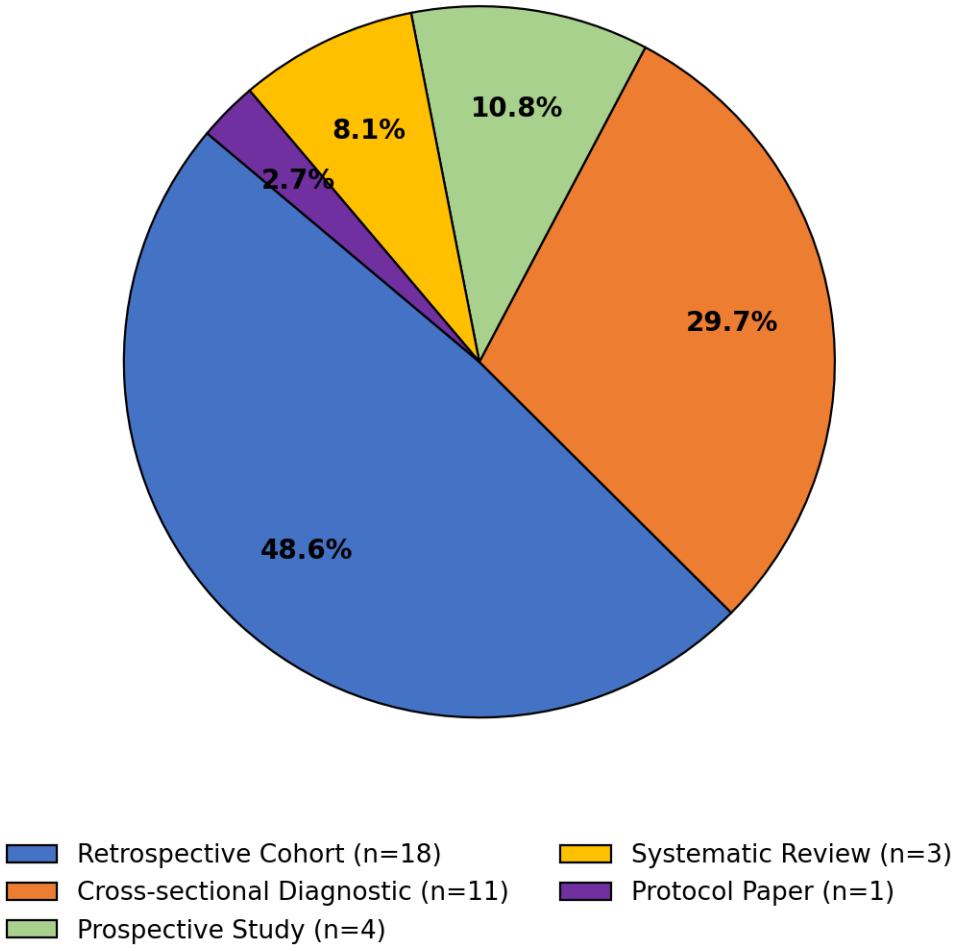


Fig 4. Distribution of study designs across the 37 included studies. Retrospective cohort studies constituted the most common design (n=18, 48.6%), followed by cross-sectional diagnostic studies (n=11, 29.7%), prospective studies (n=4, 10.8%), systematic reviews (n=3, 8.1%), and protocol papers (n=1, 2.7%).

3.1.2. Deep Learning Approaches

Deep learning constituted the most frequently employed methodology, featuring in 24 of 37 included studies (64.9%), with CNNs the dominant architecture applied to cytology and colposcopy image classification. Yuan et al.[26] developed a three-model deep learning diagnostic system for colposcopy images, achieving 84.1% accuracy, 85.4% sensitivity, 82.6% specificity, and an AUC of 0.93 by incorporating saline, acetic acid, and iodine-stained images alongside clinical data. Transfer learning using pretrained architectures such as VGG16, AlexNet, ResNet, and InceptionNet was widely employed to offset the scarcity of labelled medical image datasets: Aziz-ur-Rehman et al.[9] applied VGG16-based transfer learning to the Cervical Cells and Herlev databases, reporting two-class accuracies of 98.8-99.6% depending on classifier head, while Mathivanan et al.[10] fused AlexNet, ResNet, and InceptionNet features with classical classifiers on the SIPaKMeD dataset, reporting a top accuracy of 98.08% using a ResNet152 pipeline. Aurelia et al.[27] further demonstrated the complementary combination of CNN and SVM classifiers on standard cytology benchmarks, and Alsubai et al.[28] applied a privacy-preserving CNN pipeline to Pap smear images, addressing data-governance concerns increasingly relevant to multi-institutional deployment.

Point-of-care deep learning systems have demonstrated feasibility in resource-limited settings. Holmstrom et al.[7] validated a cloud-based digital cytology system in 740 HIV-positive women in Kenya, reporting sensitivities of 95.7-100%, specificities of 78.4-84.7%, and AUCs of 0.94-0.96 for atypia detection, while Wentzensen et al.[5] showed that deep learning automation of dual-stain cytology achieved substantially higher specificity than conventional Pap smear at equivalent sensitivity, reducing unnecessary colposcopy referrals. Li et al.[29] applied U-Net++ segmentation combined with RepVGG classification to 848

colposcopic images, reporting 83.01% overall accuracy, and Manzano Chaya et al.[30] confirmed the practical applicability of automated cytology systems in a university-population observational study. Specialised architectures have also targeted specific subtasks: Pang et al.[31] introduced PGCC-Net, validated on 20,936 cervical pathology images with state-of-the-art detection performance; Rasheed et al.[32] developed an improved UNet for cervical cell nucleus segmentation; Muksimova et al.[33] proposed RL-Cervix.Net, a lightweight reinforcement-learning classifier later extended to a novelty-classification variant[24]; Kalbhor et al.[34] proposed the CerviCell detector for cancerous-cell identification; Hut et al.[35] evaluated a whole-slide CNN classifier for liquid-based cytology; and Mondal et al.[36] combined deep learning with classical machine learning to improve benchmark performance.

3.1.3. Classical Machine Learning Approaches

Classical machine learning, including random forest, support vector machines (SVM), gradient boosting, and ensemble methods, featured in 8 of 37 included studies (21.6%), predominantly applied to structured clinical and genomic data rather than image analysis. Xiao et al.[3] conducted the largest diagnostic prediction study in this review, applying a stacking ensemble incorporating HPV genotype data, cytology results, and age across a multicentre cohort of 21,720 women; genotype inclusion raised the AUROC for CIN3+ prediction to 0.87, with 80.1% sensitivity and 83.4% specificity. Abdoh et al.[37] applied random forest with SMOTE and feature reduction to address class imbalance in risk-factor datasets, and Al Mudawi and Alazeb[38] compared multiple algorithms on a similar risk-factor dataset. Wittenborn et al.[6] employed random forest in a retrospective cohort of 334 high-risk HPV-positive patients, confirming colposcopy result and HPV PCR category as the most influential variables, while He et al.[39] and Pan et al.[40] developed and validated machine learning models for CIN grade II+ prediction and three-dimensional precancerous lesion screening, respectively. Chen et al.[25] developed a multicentre machine learning model based on HPV genotyping results, demonstrating generalisability across clinical populations. Explainable AI (XAI) approaches have begun to address the interpretability challenge in this subfield: Kafle et al.[41] applied SHAP and LIME to multiple classifiers for cervical cancer prediction, emphasising the importance of model transparency for clinician trust and regulatory acceptance.

3.1.4. Multimodal Integration Approaches

Five included studies (13.5%) employed multimodal integration, combining image data with clinical variables, HPV genotype information, or multiple imaging modalities. Miyagi et al.[12] constructed a CNN concatenated with an HPV-type tensor for classification of uterine cervical squamous epithelial lesions, achieving 94.1% accuracy, 95.6% sensitivity, 83.3% specificity, and an AUC of 0.963; incorporation of HPV-type information alongside colposcopy images outperformed both image-only AI and the expert oncologists reported in the same study. Dehingia et al.[11] proposed a cross-modal framework integrating colposcopy images, cytology results, and HPV test outcomes for screening and triage. Ahmed et al.[42] assembled a multi-institutional dataset of 9,462 women and 17,013 images, finding that combining the top-performing deep neural network with HPV type information yielded an AUC of 0.89, and Mukku and Thomas[43] proposed a hybrid decision-fusion multimodal ensemble reporting 96.16% accuracy. Baber et al.[44] similarly integrated machine learning with combined Pap smear and HPV screening data to strengthen prediction and prevention pathways. Elsewhere in the primary literature, Wang et al.[45] likewise demonstrated that AI can achieve precision diagnosis of cytology grades directly comparable to expert pathologists, reinforcing the trajectory toward high-fidelity, deployment-ready models.

3.1.5. Performance Metrics Summary

Performance metrics varied considerably across studies, reflecting differences in task definition, dataset characteristics, and model architecture. AUC values ranged from 0.85 for classical ensemble models on genotype data[3] to 0.96 for point-of-care deep learning cytology systems[7]. Accuracy ranged from 83.01%[29] to 99.99%[46], sensitivity from 80.1% to 100%, and specificity from 67.8% to 99%. A recent meta-analysis by Liu et al.[47] reported pooled AUC consistently exceeding 0.90, a finding corroborated by Takahashi et al.[19] across colposcopy-specific AI studies. Full per-study performance metrics for all 37 included studies are provided in Supplementary Table S2; Table 1 consolidates performance ranges by methodology category.

Deep learning architectures accounted for nearly two-thirds of all included studies and demonstrated the widest accuracy range, 83.0-100%, and an AUC range of 0.89-0.96, reflecting the diversity of tasks and dataset sizes addressed by these approaches. Classical machine learning models reported a narrower but consistently high accuracy range of 97.6-100%, with AUC values from 0.85-0.90, attributable in part to their application to structured tabular rather than image data. Multimodal integration approaches, despite the smallest study count, achieved the highest maximum AUC observed in this review (0.963), suggesting that combining image-based features with HPV genotype or clinical variables yields the most discriminative diagnostic performance among the three methodological categories reviewed (Figure 5).

Table 1. Consolidated Performance Summary by AI Methodology Category

AI Methodology Category	No. of Studies (n)	Accuracy Range (%)	Sensitivity Range (%)	Specificity Range (%)	Representative AUC Range
Deep Learning (CNN / Transfer Learning)	24 (64.9%)	83.01 - 99.98	84 - 100	78.4 - 99.0	0.89 - 0.96
Classical Machine Learning (RF, SVM, Ensemble)	8 (21.6%)	97.6 - 100	80.1 - 96.67	81.0 - 100	0.85 - NR
Multimodal Integration (Image + Clinical / Molecular)	5 (13.5%)	94.1 - 96.16	95.6 - NR	83.3 - NR	0.89 - 0.963
Narrative Reviews / Meta-analyses	3 (contextual)	NR	Pooled 92.1%	Pooled 90.0%	>0.90 (pooled)
Total / Overall Range	37	83.01 - 99.99	80.1 - 100	67.8 - 99.0	0.85 - 0.96

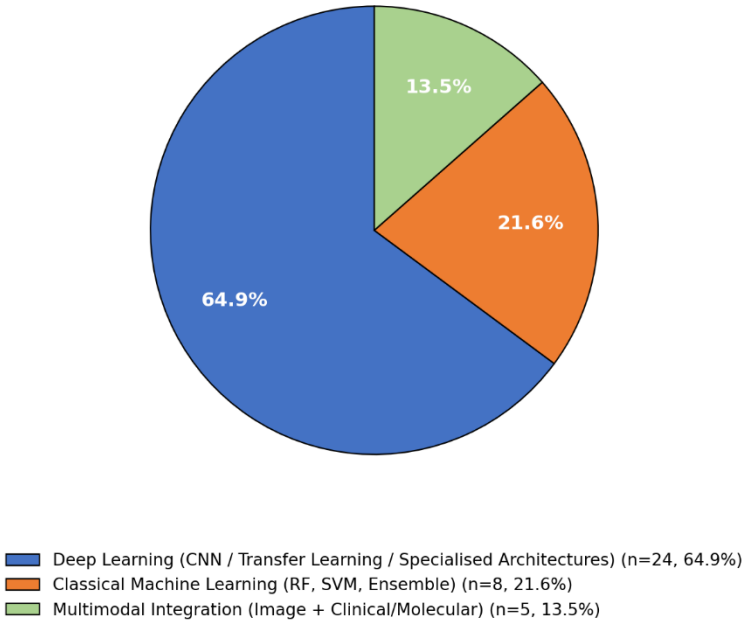


Fig 5. AI methodology distribution and accuracy comparison across the 37 included studies. Left: distribution by methodology category, deep learning (n=24, 64.9%), classical machine learning (n=8, 21.6%), and multimodal (n=5, 13.5%). Right: reported accuracy ranges by category, deep learning 83.01-99.98%, classical machine learning 97.6-100%, and multimodal 94.1-96.16%.

3.2 Milestone Studies and Technological Breakthroughs

Several studies within the included literature represent identifiable milestones, either through scale, deployment context, or methodological novelty. Table 3 consolidates the six milestones judged most consequential for the field's trajectory toward HPV genotype-integrated, clinically deployable AI.

Taken together, these milestones trace a coherent trajectory: from image-only CNN classification, through transfer learning that reduced data requirements, to multimodal genotype fusion that now defines the performance ceiling of the field, with point-of-care deployment in Kenya[7] and explainability tooling[41] marking the two translational frontiers that remain least mature.

3.3 Research Gaps, Bottlenecks, and Emerging Directions

Beyond individual study limitations, four cross-cutting gaps and bottlenecks recur across the included literature and are consolidated in Table 4, together with the emerging direction each motivates.

Table 2. Principal Datasets Used Across Included Studies

Dataset Name	Data Type	Sample Size / Images	Access	Representative Studies Using It
SIPaKMeD	Cervical cytology cell images	4,049 images	Public	[10], [28]
Herlev Pap Smear	Digitised Pap smear images	917 images	Public	[9], [27]
PAVE Multi-Country Cohort	Colposcopy + HPV VIA images	Multi-country, ongoing	Research	[48]
Multi-centre China Cohort	Colposcopy + HPV genotype records	21,720 women	Private	[3]
Multi-institutional Cervix Image DB	Colposcopy images	17,013 images / 9,462 women	Private	[42]
PGCC-Net Pathology Dataset	Whole-slide pathology images	20,936 images	Private	[31]
Kenya HIV+ Pap Smear Dataset	Digital Pap cytology slides	740 women	Private	[7]
Japan Colposcopy + HPV Dataset	Colposcopy + HPV type records	NR	Private	[12]
Cervical Cancer Risk Factor DB	Structured clinical features	Variable	Public / Private	[37], [38]
China CIN2+ Prediction Cohort	Clinical + HPV records	Cross-sectional population	Private	[39]

Table 3. Milestone Studies and Technological Breakthroughs in AI Based HPV Related Cervical Cancer Diagnosis

Study	Milestone / Breakthrough	Key Metric
Miyagi et al. [12]	First multimodal fusion of colposcopy images with an HPV-type tensor	Accuracy 94.1%, AUC 0.963 (highest ceiling performance in this review)
Xiao et al. [3]	Largest diagnostic cohort combining genotype, cytology, and age data	n=21,720; AUROC for CIN3+ raised to 0.87 with genotype inclusion
Holmstrom et al. [7]	First large-scale point-of-care deployment in a resource-limited setting	n=740 (Kenya); AUC 0.94-0.96
Wentzensen et al. [5]	Automation of dual-stain cytology reducing unnecessary colposcopy referral	Higher specificity at equivalent sensitivity to conventional Pap smear
Pang et al. [31]	Largest image dataset among included studies; state-of-the-art detection	n=20,936 cervical pathology images
Kafle et al. [41]	First application of SHAP/LIME explainability in this domain	Established an interpretability baseline for clinician trust

Table 4. Research Gaps, Bottlenecks, and Recommended Future Directions

Gap / Bottleneck	Supporting Evidence	Recommended Future Direction
Reliance on public benchmark datasets (SIPaKMeD, Herlev) limits real-world generalisability	Ahmed et al. [42]	External validation on independent, prospective, multi-site cohorts
Heterogeneous performance-reporting metrics preclude cross-study pooling	Li [49]	Adoption of a standardised performance-reporting framework
Geographic concentration in China, the United States, and Europe (73.0% of included studies)	Section 3.1.1 findings	Multi-site studies prioritising low- and middle-income settings; PAVE-style initiatives [48]
Limited interpretability of black-box deep learning models	Kafle et al. [41]	Mandated integration of XAI tools (SHAP, LIME) into diagnostic pipelines
Predominance of retrospective designs (78.4%) limits causal and real-world inference	Section 3.1.1 findings	Prospective, multi-site validation trials
Absence of an AI-specific regulatory pathway for genotype-fused diagnostic tools	Section 3.3 discussion	Development of regulatory frameworks modelled on emerging FDA / EMA / TGA approaches

A substantial proportion of included studies relied on public benchmark datasets such as SIPaKMeD and Herlev, which may not reflect real-world clinical populations; external validation on independent prospective cohorts is essential before clinical deployment [42]. Significant heterogeneity in reported performance metrics also makes direct cross-study comparison difficult, and standardisation of evaluation frameworks is urgently needed [49]. Most studies were conducted in high-resource settings, limiting generalisability to low- and middle-income countries [13], and interpretability remains a live concern, since black-box deep learning models may face clinician resistance absent transparent explanations; the application of XAI tools such as SHAP and LIME represents one promising response [41]. Future research should prioritise prospective clinical validation, federated learning, integration into existing screening workflows, and establishment of regulatory pathways; the convergence of AI with molecular diagnostics may further sharpen the precision of HPV-related cervical cancer diagnosis [3], [47]. This direction aligns with the broader PAVE study initiative [48], which aims to validate AI-based visual evaluation as a triage tool following self-sampled HPV testing in low- and middle-income countries. Zhang et al. [50] similarly identified performance variability and clinical-implementation barriers as key areas requiring attention.

3.4 Summary of Key Findings

Artificial intelligence demonstrates substantial and growing potential to enhance the diagnosis of HPV-related cervical cancer across multiple modalities and clinical settings. Deep learning approaches applied to cytology and colposcopy images consistently achieve AUC values of 0.87-0.96, with performance comparable to or exceeding that of trained clinicians in several studies [12], [7]. The integration of HPV genotype data with image-based AI models represents a particularly promising strategy, and point-of-care deployment in low-resource settings [7] demonstrates that this promise is not confined to high-income health systems.

4 Conclusion

This review contributes to the literature in a way that the ten existing reviews of AI in cervical cancer surveyed in Section 1 do not: it isolates HPV genotype-linked diagnostic evidence as a distinct analytic unit, applies dual formal risk-of-bias appraisal (QUADAS-2, PROBAST) to every included study, and reports methodology-stratified performance bands rather than aggregate figures. The resulting synthesis shows, quantitatively, that multimodal genotype-fusion models outperform image-only approaches at the upper end of the performance distribution (AUC up to 0.963 versus 0.96 for the best image-only system), a comparison not previously isolated in the review literature.

This review has five limitations. First, the search was restricted to four databases (SciSpace, PubMed/MEDLINE, Google Scholar, ArXiv), selected as described in Section 2.1; Scopus and Web of Science were not searched directly, and indexing gaps relative to these and other sources such as Embase or the Cochrane Library may have excluded eligible studies. Second, the formal search was conducted in May 2025; although a supplementary, non-systematic check was performed in July 2026 (Section 2.1), this was not a full re-screening, and studies published after May 2025 meeting the eligibility criteria may not be reflected in the synthesis. Third, the included evidence base is dominated by retrospective and cross-sectional designs (78.3%), with prospective studies comprising only 10.8%, limiting causal and real-world performance inference. Fourth, heterogeneity in reported metrics across studies precluded formal meta-analytic pooling within this review; performance ranges are reported narratively rather than as pooled point estimates. Fifth, geographic concentration of included studies in China, the United States, and Europe (73.0% combined) constrains the generalisability of the synthesised findings to under-represented low- and middle-income settings.

Based on the evidence synthesised in this review, six recommendations are proposed for researchers, clinicians, and policymakers, organised from nearer-term to longer-term priorities. First, near term, future studies should prioritise prospective, multi-site validation of AI diagnostic tools in geographically and demographically diverse populations, particularly in low- and middle-income countries. Second, standardised reporting frameworks for AI performance in cervical cancer diagnosis should be developed and adopted universally to enable cross-study comparison. Third, medium term, federated learning approaches should be explored to build robust, generalisable models without requiring centralised sharing of sensitive patient data. Fourth, the integration of XAI methods into AI diagnostic systems should be pursued as standard practice to support clinician trust and facilitate regulatory review. Fifth, longer term, regulatory bodies should develop AI-specific pathways for medical device approval, drawing on frameworks already emerging in Australia, the United States, and the European Union. Sixth, investment in digital infrastructure, including reliable connectivity, digital health records, and clinical staff training, is a prerequisite for effective AI deployment in resource-constrained settings. AI-based cervical cancer screening tools hold promises for low- and middle-income countries, where they may substantially reduce the burden of a largely preventable cancer.

5 Acknowledgements

The authors declare no funding sources for this review. The authors thank the editors and anonymous reviewers for their constructive contributions.

6 Declaration of Interest

Declarations of interest: none.

References

1) Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer Journal for Clinicians*. 2021;71(3):209–249. https://doi.org/10.3322/caac.21660.
2) Katiyar K, Jahan N, Uddin G. A review on HR HPV: An increasing cause of concern for cervical carcinoma. *Uttar Pradesh Journal of Zoology*. 2024;45(10):4053. https://doi.org/10.56557/upjoz/2024/v45i104053.
3) Xiao T, Wang C, Yang M, Yang J, Xu X, Shen L, et al. Use of virus genotypes in machine learning diagnostic prediction models for cervical cancer in women with high-risk human papillomavirus infection. *JAMA Network Open*. 2023;6(8):e2326890. https://doi.org/10.1001/jamanetworkopen.2023.26890.
4) Long NP, Jung KH, Yoon SJ, Anh NH, Nghi TD. Systematic assessment of cervical cancer initiation and progression uncovers genetic panels for deep learning based early diagnosis. *Oncotarget*. 2017;8(65):109436–109456. https://doi.org/10.18632/oncotarget.22689.
5) Wentzensen N, Lahrmann B, Clarke MA, Kinney W, Tokugawa D, Poitras N, et al. Accuracy and efficiency of deep learning-based automation of dual stain cytology in cervical cancer screening. *Journal of the National Cancer Institute*. 2021;113(1):72–79. https://doi.org/10.1093/jnci/djaa066.
6) Wittenborn J, Kupec T, Iborra S, Najjari L, Kennes LN, Stickeler E. CIN2+ detection in high-risk HPV patients with no or minor cervical cytologic abnormalities: a clinical approach validated by machine learning. *Archives of Gynecology and Obstetrics*. 2023;307:881–890. https://doi.org/10.1007/s00404-023-06953-6.
7) Holmstrom O, Linder N, Kaingu H, Mbuuko N, Mbete J, Kinyua F, et al. Point of care digital cytology with artificial intelligence for cervical cancer screening in a resource limited setting. *JAMA Network Open*. 2021;4(3):e211740. https://doi.org/10.1001/jamanetworkopen.2021.1740.
8) Ali M, Sarwar A, Sharma V, Suri J. Artificial neural network-based screening of cervical cancer using a hierarchical modular neural network architecture (HMNNA) and novel benchmark uterine cervix cancer dataset. *Neural Computing and Applications*. 2019;31:2979–2993. https://doi.org/10.1007/s00521-017-3246-7.
9) ur Rehman A, Ali N, Taj IA, Sajid M, Karimov KS. An automatic mass screening system for cervical cancer detection based on convolutional neural network. *Mathematical Problems in Engineering*. 2020;2020:4864835. https://doi.org/10.1155/2020/4864835.
10) Mathivanan SK, Francis D, Srinivasan S, Murugesan P, Rajendran S, Jayagopal P, et al. Enhancing cervical cancer detection and robust classification through a fusion of deep learning models. *Scientific Reports*. 2024;14:10834. https://doi.org/10.1038/s41598-024-61063-w.
11) Dehingia N, Raj A, Bhatt S, Dey S, Bhatt S. Deep learning based cervical screening by the cross-modal integration of colposcopy, cytology, and HPV test. *International Journal of Medical Informatics*. 2022;157:104675. https://doi.org/10.1016/j.ijmedinf.2021.104675.
12) Miyagi Y, Takehara K, Nagayasu Y, Miyake T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncology Letters*. 2020;19(2):1602–1610. https://doi.org/10.3892/ol.2019.11214.
13) Pawar M, Pingale S, Kadam K, Jadhav A, Kaul-Ghanekar R. Artificial intelligence-driven screening, early diagnosis, and treatment strategies for cervical cancer: An overview. *Infectious Agents and Cancer*. 2026;21(1):1. https://doi.org/10.1186/s13027-025-00716-5.
14) Tulimilli SV, Karnik M, Cockroft S, Bettadapura ADS, Nataraj SM, Mohideen HS, et al. Recent advancements in the application of artificial intelligence-based approaches for screening, diagnosis, prognosis and treatment of cervical cancer. *Oncology Reviews*. 2026;20:1760156. https://doi.org/10.3389/or.2026.1760156.
15) Silva ALD, Nicolli AC, Junior MLCA, Costa LC, et al. Artificial intelligence use in the routine of cervical-vaginal cytology: A systematic review. *Surgical and Experimental Pathology*. 2026;9:8. https://doi.org/10.1186/s42047-026-00222-x.
16) Kondo TS, Ngondya D, Rusheke H. Application of artificial intelligence in cervical cancer diagnosis using risk factors: A systematic review. *Telematics and Informatics Reports*. 2025;20:100250. https://doi.org/10.1016/j.teler.2025.100250.
17) Giansanti D, Lastrucci A, Pirrera A, Villani S, Carico E, Giarnieri E. AI in cervical cancer cytology diagnostics: A narrative review of cutting-edge studies. *Bioengineering*. 2025;12:769. https://doi.org/10.3390/bioengineering12070769.
18) Priya S, Panigrahi SK, Bansal V, et al. Artificial intelligence colposcopy models for cervical cancer screening and diagnosis: A systematic review and meta-analysis. *Journal of Obstetrics and Gynecology of India*. 2026. https://doi.org/10.1007/s13224-026-02352-4.
19) Takahashi Y, Kobayashi Y, Sakurai K, Matsuoka R, Akatsuka J. A systematic review of the application of artificial intelligence in colposcopy: Diagnostic accuracy for cervical intraepithelial neoplasia and cervical cancer. *Clinical Medicine Insights: Oncology*. 2025;19. https://doi.org/10.1177/11795549251374908.
20) Dellino M, Cerbone M, d'Amati A, Bochicchio M, Lagana AS, Etrusco A, et al. Artificial intelligence in cervical cancer screening: Opportunities and challenges. *AI*. 2024;5:2984–3000. https://doi.org/10.3390/ai5040144.
21) William W, Ware A. Leveraging AI and data science across the cervical cancer care continuum in developing economies. *Discover Artificial Intelligence*. 2025;5:370. https://doi.org/10.1007/s44163-025-00648-4.
22) Wang X, Wang Q, Ding G, Wang J, Tang Y, Feng Y. Artificial intelligence for colposcopic and cytological image analysis in early cervical cancer detection. *iScience*. 2026;29(2):114627. https://doi.org/10.1016/j.isci.2026.114627.
23) Nguyen HD, Le HD, Nguyen CN. A review of deep learning applications in cervical cancer cell diagnosis and screening. *TNU Journal of Science and Technology*. 2025;230(01). https://doi.org/10.34238/tnu-jst.12764.
24) Muksimova S, Umirzakova S, Shoraimov S, Baltayev A, Cho YI. Novelty classification model use in reinforcement learning for cervical cancer. *Cancers*. 2024;16(22):3782. https://doi.org/10.3390/cancers16223782.
25) Chen Z, Zeng Y, Luo L, Li J, Zheng L, Xie Y, et al. Multicenter validated machine learning model for cervical cancer based on human papillomavirus genotyping results. *International Journal of Gynecological Cancer*. 2025;35(1):102855. https://doi.org/10.1016/j.ijgc.2025.102855.

26) Yuan C, Yao Y, Cheng B, Cheng Y, Li Y, Li Y, et al. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific Reports*. 2020;10(1):11639. https://doi.org/10.1038/s41598-020-68252-3.
27) Aurelia JE, Rustam Z, Wirasati I. Cervical cancer classification using convolutional neural network-support vector machine. *Telkomnika*. 2021;19(5):1605–1611. https://doi.org/10.12928/telkomnika.v19i5.20406.
28) Alsubai S, Alqahtani A, Sha M, Shaukat A, Cao Y, Umer M. Privacy preserved cervical cancer detection using convolutional neural networks applied to Pap smear images. *Computational and Mathematical Methods in Medicine*. 2023;2023:9676206. https://doi.org/10.1155/2023/9676206.
29) Li B, Chen L, Sun C, Zhang Y, Wang Q, Zhao X. Leveraging deep learning for early detection of cervical cancer and dysplasia in China using U-NET++ and RepVGG networks. *JMIR Preprints*. 2025. https://doi.org/10.2196/preprints.70825.
30) Chaya LM, Herrera AM, Ayala IG. Use of artificial intelligence in the diagnosis of alterations in cervical cytology: A university population based observational study. *Biomedica*. 2025;45(1). https://doi.org/10.7705/biomedica.7651.
31) Pang Z, Ma X, Jiang H, Yu H. Cells grouping detection and confusing labels correction on cervical pathology images. *Bioengineering*. 2024;12(1):23. https://doi.org/10.3390/bioengineering12010023.
32) Rasheed S, Shirazi SH, Umar AI, Shahzad M, Yousaf MH. Cervical cell nucleus segmentation through an improved UNet architecture. *PLoS One*. 2023;18(3):e0283568. https://doi.org/10.1371/journal.pone.0283568.
33) Muksimova S, Umirzakova S, Baltayev A, Cho YI. RL Cervix.Net: A hybrid lightweight model integrating reinforcement learning for cervical cell classification. *Diagnostics*. 2025;15(3):364. https://doi.org/10.3390/diagnostics15030364.
34) Kalbhor M, Shinde S, Wajire P, Jude A. CerviCell detector: An object detection approach for identifying the cancerous cells in Pap smear images of cervical cancer. *Heliyon*. 2023;9(11):e22324. https://doi.org/10.1016/j.heliyon.2023.e22324.
35) Hut I, Jeftic B, Dragicevic A. Computer aided diagnostic system for whole slide image of liquid based cervical cytology sample classification using convolutional neural network. *Contemporary Materials*. 2022;13(2):169–179. https://doi.org/10.7251/comen2202169h.
36) Mondal S, Chatterjee A, Gourisaria MK, Sahni M, Castro EL. Cervical cancer classification using a novel hybrid approach. *Frontiers in Oncology*. 2025;15:1703772. https://doi.org/10.3389/fonc.2025.1703772.
37) Abdoh SF, Rizka MA, Maghraby FA. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. *IEEE Access*. 2018;6:59475–59485. https://doi.org/10.1109/ACCESS.2018.2874063.
38) Mudawi NA, Alazeb A. A model for predicting cervical cancer using machine learning algorithms. *Sensors*. 2022;22(11):4132. https://doi.org/10.3390/s22114132.
39) He Y, Liu X, Wang J, Zhang L, Chen S. Construction and validation of a model for predicting cervical intraepithelial neoplasia grade II+: A cross-sectional population study via machine learning. *International Journal of Women's Health*. 2025;17:1183–1195. https://doi.org/10.2147/IJWH.S554552.
40) Pan Y, Shen Z, He Y, Li J, Zhang M. Application of three-dimensional coding network in screening and diagnosis of cervical precancerous lesions. *Frontiers in Computational Intelligence Systems*. 2024;6(1). https://doi.org/10.54097/mi3vm0yb.
41) Kafle S, Paudel P, Bhandari M. Beyond accuracy and classification: XAI driven interpretability in cervical cancer. *INJET*. 2025;3(1). https://doi.org/10.3126/injet.v3i1.87023.
42) Ahmed SR, Befano B, Lemay A, Egemen D, Rodriguez AC, Angara S, et al. Reproducible and clinically translatable deep neural networks for cervical screening. *Scientific Reports*. 2023;13(1):21772. https://doi.org/10.1038/s41598-023-48721-1.
43) Mukku L, Thomas J. Hybrid decision fusion based multimodal ensemble framework for cervical cancer detection. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2023;9(1). https://doi.org/10.32628/cseit239018.
44) Baber R, Latif N, Raza K, Iqbal S, Javed A. Utilizing AI to predict and prevent cervical cancer: Integrating machine learning with Pap smear and HPV screening. *Journal of Neonatal Surgery*. 2025;13(1). https://doi.org/10.63682/jns.v13i1.7438.
45) Wang J, Yu Y, Tan Y, et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nature Communications*. 2024;15:4369. https://doi.org/10.1038/s41467-024-48705-3.
46) Munshi RM. Novel ensemble learning approach with SVM imputed ADASYN features for enhanced cervical cancer prediction. *PLoS One*. 2024;19(1):e0296107. https://doi.org/10.1371/journal.pone.0296107.
47) Liu Y, Liu J, Su X, Chu X, Xia J. Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: a systematic review and meta-analysis. *EClinicalMedicine*. 2025;79:102992. https://doi.org/10.1016/j.eclinm.2024.102992.
48) de Sanjose S, Perkins RB, Campos N, Inturrisi F, Egemen D, Befano B, et al. Design of the HPV automated visual evaluation (PAVE) study: Validating a novel cervical screening strategy. *eLife*. 2024;12:e91469. https://doi.org/10.7554/eLife.91469.
49) Li D. Artificial intelligence applications in cervical cancer: Current progress and future prospects. *Proceedings of the 2025 International Conference on Computational Intelligence Systems*. 2025. https://doi.org/10.1145/3757749.3757768.
50) Zhang Y, Yuan L, Chen X. Transforming cervical cancer pathological diagnosis through artificial intelligence: progress, performance, and barriers to clinical implementation. *Frontiers in Oncology*. 2025;15:1716018. https://doi.org/10.3389/fonc.2025.1716018.