

Artificial Intelligence-Based Early Detection of Cervical Cancer Using Pap Smear Images: A Comprehensive Review

Nandita M B

B.E. Biomedical Engineering
nanditamahesh1007@gmail.com

Abstract

Cervical cancer remains an important global health problem despite being largely preventable through human papillomavirus (HPV) vaccination, screening, and appropriate treatment of precancerous lesions. Pap smear cytology is an established screening approach, but manual interpretation of cervical cytology images is labor-intensive and may be affected by image quality, cellular overlap, specimen preparation, and inter-observer variability. Recent advances in artificial intelligence (AI), machine learning (ML), and deep learning (DL) have created new opportunities for computer-assisted analysis of Pap smear images. This comprehensive review examines the development of AI-based approaches for cervical cytology, with particular emphasis on image preprocessing, segmentation, feature extraction, convolutional neural networks (CNNs), transfer learning, attention mechanisms, hybrid CNN-transformer models, and ensemble methods. Publicly available datasets such as Herlev and SIPaKMeD have played a major role in benchmarking algorithms. Published studies have reported high performance on selected benchmark datasets, including binary and multiclass classification tasks. For example, DeepPap reported 98.3% classification accuracy and an area under the curve of 0.99 on the Herlev dataset, while DeepCervix reported accuracies of 99.85%, 99.38%, and 99.14% for different SIPaKMeD tasks and 98.32% for binary classification on Herlev. More recent work has explored feature fusion, vision transformers, causal approaches, and joint segmentation-classification frameworks. However, reported performance across studies cannot be directly compared because of differences in datasets, preprocessing, class definitions, validation strategies, and outcome metrics. External validation, dataset diversity, explainability, calibration, and prospective clinical evaluation remain important requirements before routine clinical deployment. AI should therefore be considered a decision-support technology rather than a replacement for qualified cytopathology and clinical assessment. Overall, AI-assisted Pap smear analysis is a promising research direction for improving the efficiency and consistency of cervical cytology screening, particularly when developed and validated using representative clinical data.

Keywords

Cervical cancer; Pap smear; cervical cytology; artificial intelligence; machine learning; deep learning; convolutional neural networks; medical image analysis; computer-aided diagnosis; early detection; transfer learning; image classification; explainable AI.

1. Introduction

Cervical cancer is a major preventable malignancy and continues to produce substantial morbidity and mortality worldwide. The disease is strongly associated with persistent infection by high-risk human papillomavirus (HPV), and effective primary and secondary prevention strategies can substantially reduce its burden [1–3]. The World Health Organization (WHO) estimates that approximately 604,000 new cases and 280,000 deaths occurred globally in 2024, with the greatest burden concentrated in low- and middle-income countries [1]. These disparities reflect differences in vaccination coverage, access to screening, diagnostic capacity, treatment, and health-system infrastructure.

Screening is central to cervical cancer prevention because precancerous changes can be detected before invasive disease develops. Cytological examination of cervical cells, commonly known as the Pap smear or Pap test, has historically been an important screening method. Cytologists evaluate cellular morphology, including nuclear size, nuclear-to-cytoplasmic ratio, chromatin pattern, cell shape, staining characteristics, and other abnormalities. Although cytology can be highly informative, conventional microscopy requires substantial human expertise and time. Dense cellular clusters, overlapping cells, staining variation, artifacts, and poor-quality specimens can further complicate interpretation.

Computer-assisted analysis has therefore attracted increasing attention. Early image-analysis systems relied on manually engineered features, including texture, color, shape, and morphological descriptors. These approaches can be effective under controlled conditions but often depend heavily on segmentation quality and feature-engineering decisions. Deep learning has changed this paradigm by allowing neural networks to learn hierarchical representations directly from images. CNNs have become particularly prominent because they can learn spatial patterns associated with cellular morphology.

Research in cervical cytology has progressed from isolated-cell classification toward more complex tasks such as nuclei segmentation, cell detection, whole-slide-image analysis, multiclass classification, transfer learning, attention mechanisms, feature fusion, and transformer-based architectures. DeepPap demonstrated that deep convolutional networks could classify cervical cells without requiring accurate prior segmentation and reported strong performance on the Herlev benchmark [4]. DeepCervix subsequently investigated hybrid deep feature fusion on SIPaKMeD and Herlev and reported high accuracy across binary and multiclass tasks [5].

Despite promising benchmark results, an important distinction must be made between technical image-classification performance and clinical diagnostic performance. Many published models are evaluated on relatively small, curated datasets and may use random image-level train/test splitting. Images from the same patient, slide, or cluster can therefore create information leakage if related images appear in both training and testing sets. Performance can also decline when models encounter different microscopes, staining protocols, laboratories, patient populations, or disease prevalence. Consequently, clinical translation requires robust external validation and careful evaluation of generalizability.

This review summarizes the current landscape of AI-based Pap smear image analysis, identifies the major methodological approaches and datasets, compares representative published results, and discusses the barriers that must be addressed before AI systems can be safely integrated into cervical cancer screening workflows.

2. Review Objectives

The objectives of this comprehensive review are to:

- Describe the role of AI and deep learning in automated cervical cytology image analysis.
- Summarize commonly used Pap smear image datasets and their characteristics.
- Review major preprocessing, segmentation, classification, and feature-learning approaches.
- Compare representative published studies and the performance metrics they report.
- Identify methodological limitations that affect comparison and clinical generalizability.
- Discuss explainability, bias, validation, and implementation considerations for future clinical use.
- Identify promising research directions for reliable, clinically validated AI-assisted cervical screening.

3. Methodology

3.1 Review Design

This manuscript was structured as a comprehensive narrative review of published research on AI-based analysis of Pap smear and cervical cytology images. The synthesis focuses on studies that investigate image-based machine learning or deep learning for detection, segmentation, classification, or computer-assisted interpretation of cervical cells. The reporting approach was informed by principles of transparent evidence synthesis and the PRISMA 2020 framework [6]. Because an independently registered systematic-review protocol, exhaustive database screening log, and reproducible study-selection counts were not established for this manuscript, it should not be represented as a completed PRISMA systematic review.

3.2 Literature Scope

Relevant literature was identified from biomedical and scholarly sources, including PubMed-indexed research and authoritative sources from WHO and IARC. Search concepts included combinations of the terms cervical cancer, cervical cytology, Pap smear, artificial intelligence, machine learning, deep learning, convolutional neural network, image classification, segmentation, transfer learning, and computer-aided diagnosis. Priority was given to studies that reported datasets, model architectures, validation strategies, and quantitative performance measures.

3.3 Inclusion Focus

- Studies using cervical cytology, Pap smear, or related cervical-cell microscopic images.
- Studies applying ML, DL, CNN, transformer, ensemble, or related computer-vision techniques.
- Studies addressing detection, segmentation, classification, or computer-assisted interpretation.
- Studies reporting measurable outcomes such as accuracy, sensitivity, specificity, F1-score, AUC, Dice score, or mean average precision.
- Peer-reviewed studies and authoritative technical or public-health sources relevant to cervical cancer screening.

3.4 Exclusion Focus

- Studies unrelated to cervical cytology or cervical cancer screening.
- Purely clinical studies with no image-analysis or AI component.
- Non-cervical medical-image applications.

- Reports without sufficient methodological information to interpret the AI task.

4. AI-Based Cervical Cytology Workflow

AI-assisted Pap smear analysis can be represented as a pipeline consisting of image acquisition, preprocessing, segmentation or detection, feature learning, classification, post-processing, and clinical decision support. The exact pipeline varies by study and by whether the input consists of isolated cells, cell clusters, or whole-slide images.

Stage	Purpose
1. Image acquisition	Microscopic Pap smear images are captured using conventional or liquid-based cytology workflows and digital microscopy.
2. Preprocessing	Resizing, normalization, denoising, contrast adjustment, color normalization, and artifact reduction may be applied.
3. Segmentation/detection	Cell or nucleus boundaries may be identified using classical image processing, U-Net-like networks, Mask R-CNN, or other detectors.
4. Feature learning	CNNs learn spatial and morphological representations; transformer and hybrid models can capture broader contextual relationships.
5. Classification	Models classify cells into normal, abnormal, benign, precancerous-related, or other study-specific categories.
6. Evaluation	Accuracy, sensitivity, specificity, precision, recall, F1-score, AUC, Dice coefficient, and related metrics are used.
7. Clinical decision support	A validated system may prioritize suspicious images for expert review rather than replace clinical judgment.

5. Major Datasets for Cervical Cytology AI

Benchmark datasets are essential for reproducible algorithm development. Two frequently used datasets are Herlev and SIPaKMeD. SIPaKMeD contains 4,049 isolated-cell images extracted from 966 cell-cluster images and is divided into five cell categories [7]. The Herlev dataset contains 917 single-cell cervical cytology images across seven cell types, ranging from normal cells to carcinoma [8]. These datasets have supported numerous classification studies but are relatively small compared with datasets commonly used in general computer vision.

Dataset	Images	Classes	Typical Use	Key Limitation
SIPaKMeD	4,049 isolated cells from 966 clusters	5	Normal/abnormal/benign cell classification	Curated benchmark;

				limited clinical diversity
Herlev	917 single cells	7	Normal vs. abnormal and multiclass classification	Small dataset and controlled acquisition
CRIC	Varies by study	Study-specific	Cervical-cell classification	Availability and task definitions vary
Mendeley LBC	963 in commonly reported subset	4	Liquid-based cytology classification	Different labels and imaging conditions

5.1 SIPaKMeD Class Distribution

Cell type	Number of images	Broad category
Superficial-Intermediate	831	Normal
Parabasal	787	Normal
Metaplastic	793	Benign
Koilocytotic	825	Abnormal
Dyskeratotic	813	Abnormal

6. Artificial Intelligence Techniques

6.1 Conventional Machine Learning

Traditional machine-learning systems commonly rely on engineered features derived from cell morphology, color, texture, and geometry. Algorithms such as support vector machines, random forests, k-nearest neighbors, decision trees, and ensemble classifiers can then be trained using these features. Their principal limitation is dependence on the quality and completeness of the handcrafted representation. Variability in staining and segmentation can significantly affect the resulting features.

6.2 Convolutional Neural Networks

CNNs learn hierarchical image representations through convolutional filters. Lower layers typically capture local patterns such as edges and textures, while deeper layers learn more complex morphology. DeepPap demonstrated the feasibility of classifying cervical cells using deep convolutional features without requiring accurate prior segmentation [4]. CNN-based approaches remain common because of their strong performance and comparatively mature implementation ecosystem.

6.3 Transfer Learning

Transfer learning uses models pretrained on large image datasets and adapts them to cervical-cell classification. Architectures such as ResNet, VGG, DenseNet, Inception, and EfficientNet have been used as feature extractors or fine-tuned classifiers. Transfer learning can be useful when medical datasets are small, but domain differences between natural images and cytology images mean that careful fine-tuning and validation remain necessary.

6.4 Segmentation and Detection

Segmentation is important when the clinical task requires identification of individual cells or nuclei within a complex field. Hussain et al. described a fully convolutional approach for simultaneous nuclei segmentation and classification and reported strong segmentation and binary-classification performance [9]. More recent frameworks combine segmentation with downstream classification, reflecting the need to manage cell overlap and heterogeneous backgrounds.

6.5 Attention, Feature Fusion and Transformer Models

Attention mechanisms allow a model to emphasize informative regions or channels. Hybrid architectures may combine CNNs, which are effective at local feature extraction, with transformer-based modules that model broader contextual relationships. Fang et al. proposed a Deep Integrated Feature Fusion block that combined local CNN and global visual-transformer features and evaluated the approach on multiple cervical-cell datasets [10]. This direction is promising because cervical-cell morphology depends on both local nuclear details and relationships among surrounding structures.

6.6 Ensemble and Hybrid Learning

Ensemble systems combine predictions from multiple models or feature representations. The motivation is that different architectures may capture complementary information. DeepCervix, for example, used hybrid deep feature fusion and reported high classification accuracy on SIPaKMeD and Herlev [5]. Although ensemble models can improve benchmark performance, they may also increase computational complexity and make interpretation more difficult.

7. Representative Published Results

The following table summarizes representative studies identified in the literature. Values are reported as stated by the original studies and should not be interpreted as directly comparable because datasets, class definitions, validation protocols, and experimental conditions differ.

Study	Year	Dataset	Approach	Reported result	Important note
DeepPap [4]	2017	Herlev + LBC	Deep CNN	98.3% accuracy; AUC 0.99 on Herlev	Five-fold CV; benchmark dataset
DeepCervix [5]	2021	SIPaKMeD, Herlev	Hybrid deep feature fusion	SIPaKMeD: 99.85%, 99.38%, 99.14%; Herlev binary 98.32%	Different tasks/classes
Xu et al. [11]	2023	SIPaKMeD, Herlev	Improved Faster R-CNN + feature enhancement	Maximum accuracy 99.81%; overall mAP 89.4%	Detection/classification framework
Fang et al. [10]	2024	SIPaKMeD, CRIC, Herlev	CNN + visual transformer	Effective across multiple multiclass/binary tasks	Exact performance varies by task

			feature fusion		
Tan et al. [12]	2024	Pap smear datasets	Deep CNN models	High classification performance reported	Performance depends on dataset/task
CNN-based Pap smear method [13]	2025	SIPaKMeD, Herlev	Transfer-learning CNNs	96.74% WSI patches; 97.55% full-cell SIPaKMeD; 90.42% Herlev	Direct classification without segmentation
CausalCervixNet [14]	2025	SIPaKMeD, Herlev, ShUCSEIT	Causal CNN framework	99.14%, 97.31%, 99.09% accuracy respectively	Requires external validation beyond benchmarks

8. Key Findings

- AI methods consistently demonstrate the ability to learn clinically relevant cellular morphology from Pap smear images in controlled experimental settings.
- Deep-learning models generally reduce dependence on manually designed image features and can integrate color, texture, nuclear morphology, and spatial patterns.
- Segmentation remains a major challenge when cells overlap or when whole-slide images are analyzed. However, several studies demonstrate that direct classification of appropriately prepared image patches can reduce the need for perfect segmentation.
- Benchmark datasets such as SIPaKMeD and Herlev are valuable for algorithm comparison but do not fully represent the diversity of clinical laboratories, patient populations, staining protocols, and imaging systems.
- Reported accuracy values are often high, but accuracy alone can be misleading, particularly in imbalanced multiclass problems. Sensitivity, specificity, precision, recall, F1-score, AUC, calibration, and patient-level validation should be reported.
- Recent work is moving toward hybrid CNN-transformer architectures, attention mechanisms, feature fusion, ensembles, and causal or explainable approaches.
- Clinical deployment requires evidence beyond retrospective benchmark testing, including external validation, prospective evaluation, robustness testing, and assessment of workflow impact.

9. Discussion

The literature demonstrates substantial progress in automated cervical cytology. The earliest deep-learning systems established that learned representations could outperform or complement handcrafted features. Subsequent work has increasingly addressed practical limitations such as cell clustering, class imbalance, feature diversity, and the need to integrate local and global information.

One of the most important findings is the gap between benchmark performance and clinical generalizability. A model can achieve very high accuracy on a public dataset while performing less reliably

on images obtained from another laboratory. The reason is not necessarily a weakness of deep learning itself; rather, the model may learn acquisition-specific characteristics such as staining intensity, microscope characteristics, background patterns, or dataset-specific artifacts. This phenomenon emphasizes the importance of external validation.

Another major issue is the unit of evaluation. If multiple cells originate from the same smear or patient, splitting images randomly into training and testing sets can produce optimistic estimates because visually related examples may appear in both subsets. Patient-level or slide-level separation is preferable when the dataset structure allows it. Prospective validation should ideally preserve the clinical workflow and evaluate performance on previously unseen patients and laboratories.

Class imbalance is also important. SIPaKMeD has five cell categories with different sample counts, and some clinical abnormalities are inherently less frequent than normal findings. A model that performs well on common classes may still miss rare but clinically important abnormalities. Macro-averaged metrics, class-specific sensitivity, confusion matrices, and precision-recall curves can provide a more informative evaluation.

Interpretability is another requirement for clinical adoption. Pathologists and clinicians need to understand why a model flags a cell or image as suspicious. Saliency maps, attention maps, concept-based explanations, uncertainty estimates, and case-level visual evidence may improve trust. However, an explanation method should itself be validated rather than assumed to be clinically meaningful.

Finally, AI should be positioned as an assistive technology. In screening settings, a practical system could prioritize suspicious images, reduce repetitive manual review, provide quality-control alerts, or serve as a second reader. Such human-in-the-loop approaches may be more realistic than attempting to replace expert cytopathologists.

10. Challenges and Limitations

Challenge	Why it matters
Small and curated datasets	Many studies use benchmark datasets with hundreds or a few thousand images, which limits diversity and can encourage overfitting.
Data leakage	Random image-level splitting can cause related cells from the same slide or patient to appear in training and testing sets.
Domain shift	Changes in microscope, staining, preparation, camera, laboratory, and patient population can reduce model performance.
Class imbalance	Rare abnormal classes may be underrepresented, making accuracy insufficient as a standalone metric.
Lack of external validation	Many models are tested only on the dataset used for development.

Limited explainability	High-performing models may provide little information about the morphological evidence behind predictions.
Clinical endpoint mismatch	Cell-level classification is not identical to patient-level cervical cancer diagnosis or screening outcome.
Reproducibility	Differences in preprocessing, augmentation, splits, and evaluation protocols make direct comparison difficult.

11. Future Research Directions

- Develop larger, multicenter and demographically diverse cervical cytology datasets.
- Use patient-level or slide-level separation during training and testing to minimize information leakage.
- Perform external validation using independent laboratories and imaging systems.
- Report sensitivity, specificity, precision, recall, F1-score, AUC, calibration, and confidence intervals rather than accuracy alone.
- Investigate self-supervised and foundation-model approaches that can use large quantities of unlabeled cytology images.
- Develop explainable AI methods that provide clinically meaningful morphological evidence.
- Evaluate uncertainty and abstention mechanisms so the model can defer difficult cases to experts.
- Integrate image analysis with clinical and laboratory information when appropriate and ethically justified.
- Conduct prospective studies assessing whether AI reduces workload, improves turnaround time, or improves clinically relevant screening outcomes.
- Establish standardized benchmark protocols and reporting requirements for cervical cytology AI.

12. Potential Clinical Applications

- Second-reader support for cytotechnologists and pathologists.
- Prioritization of suspicious fields for expert review.
- Automated quality control for inadequate or low-quality images.
- Educational and training tools for cytology learners.
- Decision support in laboratories with limited access to specialist cytopathology expertise.
- Research platforms for quantitative analysis of cervical-cell morphology.

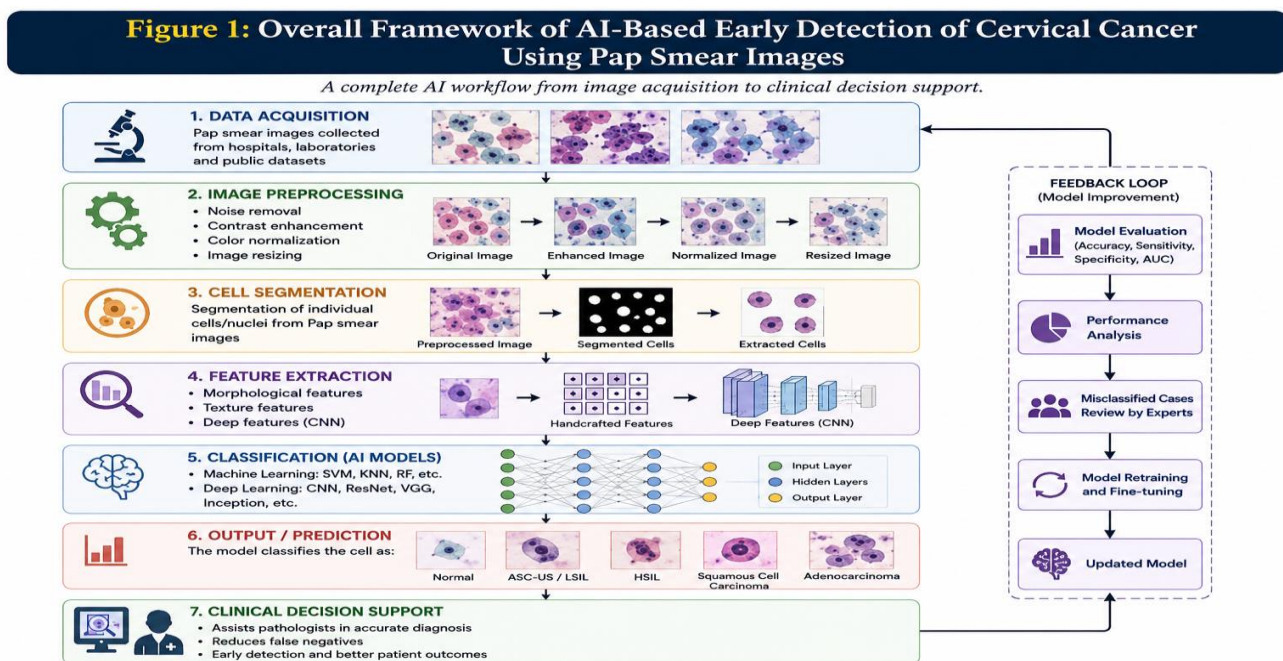
Any clinical implementation should follow applicable medical-device, data-protection, laboratory, and clinical-governance requirements. AI output should be reviewed within the established clinical workflow and should not be treated as a standalone diagnosis without appropriate validation and professional oversight.

13. Conclusion

Artificial intelligence has become an important research direction in automated cervical cytology. Studies using Pap smear images demonstrate that CNNs, transfer-learning models, segmentation networks, hybrid architectures, transformer-based systems, and ensembles can achieve strong performance on benchmark datasets. The evidence is encouraging, particularly for automated cell classification and prioritization of suspicious images. Nevertheless, high benchmark accuracy should not be interpreted as proof of clinical readiness. Dataset limitations, domain shift, data leakage, class imbalance, limited external validation, and inadequate interpretability remain significant barriers. Future studies should prioritize multicenter datasets, patient-level validation, standardized reporting, explainability, uncertainty estimation, and prospective clinical evaluation. A carefully validated human-in-the-loop AI system may ultimately improve the efficiency and consistency of cervical screening while preserving the role of trained healthcare professionals in final clinical decision-making.

Figure 1.

Overall Framework of AI-Based Early Detection of Cervical Cancer Using Pap Smear Images.



References

1. World Health Organization. Cervical cancer. WHO Fact Sheet. 3 July 2026.
2. World Health Organization. Human papillomavirus and cancer. WHO Fact Sheet. 5 March 2024.
3. World Health Organization. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention: use of dual-stain cytology to triage women after a positive test for human papillomavirus (HPV). 2nd ed. Geneva: WHO; 2024.
4. Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. IEEE? [Bibliographic details should be verified against the journal record before submission]. 2017. PMID: 28541229.

5. Rahaman MM, Li C, Yao Y, Kulwa F, Wu X, Li X, Wang Q. DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Computers in Biology and Medicine*. 2021;136:104649. doi:10.1016/j.compbimed.2021.104649.
6. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.
7. Plissiti ME, Dimitrakopoulos P, Sfikas G, Nikou C, Krikoni O, Charchanti A. SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images. *IEEE International Conference on Image Processing (ICIP)*. 2018.
8. Jantzen J, Norup J, Dounias G, Bjerregaard B. Pap-smear benchmark data for pattern classification. *Proceedings of NiSIS 2005*; 2005.
9. Hussain E, Mahanta LB, Das CR, Choudhury M, Chowdhury M. A shape context fully convolutional neural network for segmentation and classification of cervical nuclei in Pap smear images. *Artificial Intelligence in Medicine*. 2020;107:101897. doi:10.1016/j.artmed.2020.101897.
10. Fang M, Fu M, Liao B, Lei X, Wu FX. Deep integrated fusion of local and global features for cervical cell classification. *Computers in Biology and Medicine*. 2024;171:108153. doi:10.1016/j.compbimed.2024.108153.
11. Xu L, Cai F, Fu Y, Liu Q. Cervical cell classification with deep-learning algorithms. *Medical & Biological Engineering & Computing*. 2023;61(3):821–833. doi:10.1007/s11517-022-02745-3.
12. Tan SL, Selvachandran G, Ding W, Paramesran R, Kotecha K. Cervical Cancer Classification From Pap Smear Images Using Deep Convolutional Neural Network Models. *Interdisciplinary Sciences: Computational Life Sciences*. 2024;16(1):16–38. doi:10.1007/s12539-023-00589-5.
13. CNN based method for classifying cervical cancer cells in pap smear images. 2025. PMID: 40615498. Full bibliographic details should be verified from the final journal record before submission.
14. CausalCervixNet: convolutional neural networks with causal insight in cervical cancer cell classification. 2025. PMID: 40181353. Full bibliographic details should be verified from the final journal record before submission.
15. Multi-scale feature integration with enhanced cytomorph for high-accuracy cervical cytology classification. *PLOS ONE*. 2026. doi:10.1371/journal.pone.0351063.
16. World Health Organization / International Agency for Research on Cancer. Cervical cancer: global burden and prevention information. 2022–2026 data.