

AI for cervical cancer screening on whole slide images: opportunities with open-source simple tools

Laura Nonnis¹, Lorenzo Putzu², Michela Vincis¹, Stefano Guerriero³, Marco Palomba³, Alessio Rocca³, Daniela Fanni, Clara Gerosa¹, Marcello Trucas⁴

¹ Department of Medical Sciences and Public Health, University of Cagliari, Unit of Anatomical Pathology, Cittadella Universitaria, Monserrato, Italy; ² Department of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, Cagliari, Italy; ³ Department of Maternal and Child Health, University of Cagliari, Integrated Center for Medically Assisted Reproduction (PMA) and Obstetric-Gynecological Diagnostics, Cittadella Universitaria, Monserrato, Italy; ⁴ Department of Biomedical Sciences, University of Cagliari, Unit of Citomorphology, Cittadella Universitaria, Monserrato, Italy

Summary

Objective. Cervical cancer remains a major global health burden, where early detection is critical. Cytological and histological assessments aim to identify precancerous squamous intraepithelial lesions (SILs). While artificial intelligence and machine learning have shown promise, most approaches rely on cytology or are not tailored for SIL classification. The aim of this study is to develop and evaluate a weakly supervised, pixel-level machine learning framework for the histological classification of low grade and high grade SIL in whole slide images (WSIs). Specifically, we sought to assess whether an open source segmentation pipeline trained on sparsely annotated WSIs could accurately support slide-level diagnostic interpretation while minimizing annotation burden and maintaining clinical interpretability.

Methods. We propose a weakly supervised machine learning framework for classifying low grade and high grade SILs in whole-slide histological images. Using Random Forest classifiers for pixel-level segmentation, the system mimics pathologists by quantifying tissue components. Training required only sparse annotations from a limited set of WSIs, yielding millions of pixel-level samples and reducing annotation burden.

Results. Applied on a test set of 309 cervical WSIs, the system achieved over 96% concordance with expert pathologists, correctly distinguishing low grade LSIL, high grade HSIL, and normal epithelium, with only one false negative and a 7-10 false positives, depending on the used model.

Conclusions. Our approach offers accurate, interpretable, and low-cost diagnostic support, with potential for integration into routine workflows, especially in resource-limited settings.

Keywords: uterine cervical neoplasms, squamous intraepithelial lesions, histological techniques, machine learning, computer-assisted diagnosis

Introduction

Cervical cancer remains a major global health issue, especially in areas lacking organized screening programs. Early detection of squamous intraepithelial lesions (SILs) classified as low-grade (LSIL) or high-grade (HSIL) is crucial. While cervical intraepithelial neoplasia (CIN) remains common in histopathology, SIL terminology is now standard in cytology (Bethesda System) ², improving clinical communication. Histologically, LSIL (CIN1) affects the lower third of the epithelium, with koilocytosis and preserved maturation. HSIL includes CIN2 (lower two-thirds involve-

Received: October 23, 2025

Accepted: March 26, 2026

Correspondence

Lorenzo Putzu

E-mail: lorenzo.putzu@unica.it

How to cite this article: Nonnis L, Putzu L, Vincis M, et al. AI for cervical cancer screening on whole slide images: opportunities with open-source simple tools. *Pathologica* 2026;118:85-95 <https://doi.org/10.32074/1591-951X-1763>

© Copyright by Società Italiana di Anatomia Patologica e Citopatologia Diagnostica, Divisione Italiana della International Academy of Pathology



OPEN ACCESS

This is an open access journal distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license: the work can be used by mentioning the author and the license, but only for non-commercial purposes and only in the original version. For further information: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

ment, moderate atypia) and CIN3 (full-thickness atypia, high mitotic activity)¹. Cytology-based screening (e.g., Pap smear) is effective but limited by subjectivity and labor intensity, while histological images preserve tissue architecture, essential for accurate grading.

In recent years, artificial intelligence (AI) and machine learning (ML) have been increasingly explored for assisting cervical cancer screening. However, the vast majority of AI studies rely on cytology datasets, such as Herlev⁸, SiPaKMeD⁹, Mendeley LBC¹⁰, and CRIC¹¹, which offer cell-level annotations and high-resolution images, ideal for feature extraction and classification. While these datasets have enabled rapid methodological progress, they do not reflect the complexity of tissue level diagnosis, where architectural patterns and epithelial thickness are essential components of SIL grading.

Conversely, histology-based AI studies remain limited, and public datasets are scarce. To date, CAISHI¹² is the only publicly available collection of cervical histological images, but it focuses exclusively on adenocarcinoma in situ (AIS) and does not include SIL grading. In general, existing works on cervical histology typically address tasks different from LSIL/HSIL distinction:

- Li et al. (2019)¹³ applied weakly supervised learning for cervical cancer differentiation; Jiang et al. (2022)¹⁴ used a multiinstance learning approaches to classify CIN categories or cancer vs noncancer WSIs;
- Song et al. (2022)¹⁵ exploited convolutional neural networks (CNNs) for differentiating cervical vs. endometrial cancers and subtyping adenocarcinomas (ACs);
- Li Y et al. (2022)¹⁶ compared different CNNs architectures in distinguishing squamous cell carcinoma (SCC) from AC.

Computational Pathology (CPATH) merges AI with digital pathology for diagnosis and education. Recent AI systems have achieved performances comparable to human experts on narrowly defined tasks^{15,17}. Although fully supervised deep learning (DL) methods can achieve high accuracy, they require exhaustive pixel level annotations which are rarely feasible in histopathology due to the substantial time and expertise required. In cervical histology, this challenge is amplified by the need to annotate large epithelial surfaces and subtle intraepithelial variations. In addition, most AI applications in cervical cancer still rely on cytological datasets or small field-of-view images, which limits their applicability to LSIL/HSIL grading on whole slide histological specimens. Annotated histology datasets, particularly those including surgical cone samples, remain scarce. These constraints underline the need for

approaches that can operate effectively with sparse annotations while maintaining interpretability, which directly motivates the weakly supervised pixel-level framework adopted in this study. This design choice aligns with real-world constraints in routine pathology and ensures that the method remains deployable in low resource settings. Key contributions of this paper:

- we propose a semisupervised semantic segmentation method specifically designed to support pathologists in the differentiation of SIL and HSIL during lesion assessment;
- to develop and validate the method, we created a partially annotated WSI dataset, whose purpose is to demonstrate that the proposed approach does not require densely annotated slides. This shows that the method can be easily trained in other centers with minimal annotation effort, substantially reducing annotation time and costs;
- finally, we demonstrate the feasibility of integrating our method into routine research and training workflows by running it directly within QuPath. Leveraging QuPath's widespread adoption in digital pathology, this ensures accessibility, reproducibility, and suitability even for low resource settings, without the need for custom software or specialized hardware.

Materials and methods

DATASET

For our study, we exploited archived and anonymized histopathological material collected at the Gynecology and Obstetrics department of the University Hospital of Cagliari between 2021 and 2023 as part of routine clinical practice. The study concerns patients aged between 20 and 82 years with an average age of 39.88 years and was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was waived by the Institutional Review Board since the analysis involved only anonymized retrospective data from patients who had given consent at the time of surgery.

Almost 6 samples from 60 patients with a diagnosis of cervical intraepithelial neoplasia were collected, specifically 37 with a diagnosis of LSIL (CIN1) and 23 with a diagnosis of HSIL (CIN2 and CIN3), obtained after the same number of surgical interventions (30 biopsies and 30 conizations). The samples were fixed in formalin 10%, processed by dehydration in an increasing alcohol scale, clarified and embedded in paraffin. Serial sections of 3 μ m thickness were then prepared for each sample, one of which was rehy-

drated, deparaffinized and stained with the standard hematoxylin and eosin staining for histomorphology evaluation. A total of 355 slides were generated from the archived cases.

The hematoxylin and eosin stained slides were digitised at 40× magnification (0.25 µm/pixel) using a Leica Aperio GT 450 DX whole slide scanner. Digital slides were stored and processed on a MINISFORUM MiniPC UM773 Lite equipped with an AMD Ryzen 7 7735HS CPU (8 cores, 16 threads, 3.2-4.75 GHz) and 32 GB DDR5 RAM. All WSIs underwent visual quality control to verify the presence of intact epithelium, sufficient staining quality, and the absence of major artefacts. Four slides were excluded due to the complete absence of epithelial tissue, fundamental for diagnosis, resulting in a final dataset of 351 WSIs.

Three pathologists (D.F., L.N., M.V.) globally annotated all the slides that made the same diagnoses on the samples, so there was no discordance between them. Although the dataset is nearly classbalanced, as can be observed in Table I, since LSIL and HSIL share several common regions of interest at the tissue level, to promote learning of rarer HSIL-specific patterns, we applied classbalanced sampling that selected twice as many HSIL slides as LSIL slides for training. A representative subset of 42 WSIs was selected as the training set, and the remaining subset, composed of 309 WSIs was used as the testing set. This relatively small number of training images does not represent a performance-driven choice, but rather reflects realistic constraints, where large-scale manual annotation is rarely feasible. It is important to note that, since all WSIs were obtained as fully deidentified, without patient-level linkage, the train/test partition was performed at the slide level (WSI-level) via random allocation. We acknowledge that this design may allow slides from the same (unknown) patient to appear in both sets. At the same time, with only 60 patients available, a patient-level split would have further reduced inter-patient variability, increasing the risk of overfitting to patient-specific properties. The full statistics of the used dataset are reported in Table I.

It must also be noted that the training images have also been annotated at the local, pixel-level using

QuPath's manual drawing tools, providing local, point-wise (sparse) annotations of individual regions of interest. Because the strategy was intentionally sparse and focused on diagnostically meaningful areas, only a subset of regions was annotated on each slide; on average, ~50 regions per WSI were delineated and explicitly labeled as one of eight classes: HSIL, LSIL, endocervical mucosa, exocervical mucosa, immune cells, erythrocytes, stroma and ignore. Figure 1 provides a visual summary of how sparse, region-level annotations were drawn on H&E WSIs.

While HSIL and LSIL were the primary diagnostic targets, the remaining classes were annotated to enhance model performance by providing anatomical context and aiding in the exclusion of irrelevant tissue types.

This annotation strategy offers a significant advantage over most publicly available histopathology datasets, which provide image-level annotations only, our dataset also includes fine-grained, cellular-level labels, albeit in a sparse fashion. Crucially, the sparsity reflects dataset creation constraints, i.e. the substantial effort required for exhaustive pixel-level annotation in large WSIs. This enables the development and evaluation of cell-level models while remaining feasible under realistic annotation budgets.

METHOD

To support the diagnostic process from WSIs, we implemented a pixel-level semantic classification method explicitly designed to reproduce, in computational terms, the initial analytic step typically performed by pathologists. In routine practice, pathologists visually identify and compare tissue regions based on their morphological characteristics, implicitly segmenting the epithelium into interpretable components. Our method mirrors this process by assigning each pixel to one of the predefined epithelial classes through a supervised classifier, producing a structured and interpretable segmentation map of the tissue. This segmentation map is not intended as a final diagnosis, but as a structured, interpretable representation that a pathologist can directly inspect. The final LSIL/HSIL assessment therefore arises from the expert review of the segmentation maps, where the pathologist integrates quantitative and spatial information produced by the model. This human-based interpretation is part of the experimental evaluation protocol, not of the computational method itself. The central component is therefore the pixel-level segmentation module, inspired by previous work on learning-by-sampling methods for cellular component segmentation in haematological images¹⁸, adapted to histopathological tissue analysis. Figure 2 provides a conceptual

Table I. Statistics of the dataset used in terms of the number of images per partition and per diagnosis

Partition	# TOT Images	# LSIL Images	# HSIL Images
Full dataset	351	176	175
Training Set	42	14	28
Testing Set	309	162	147

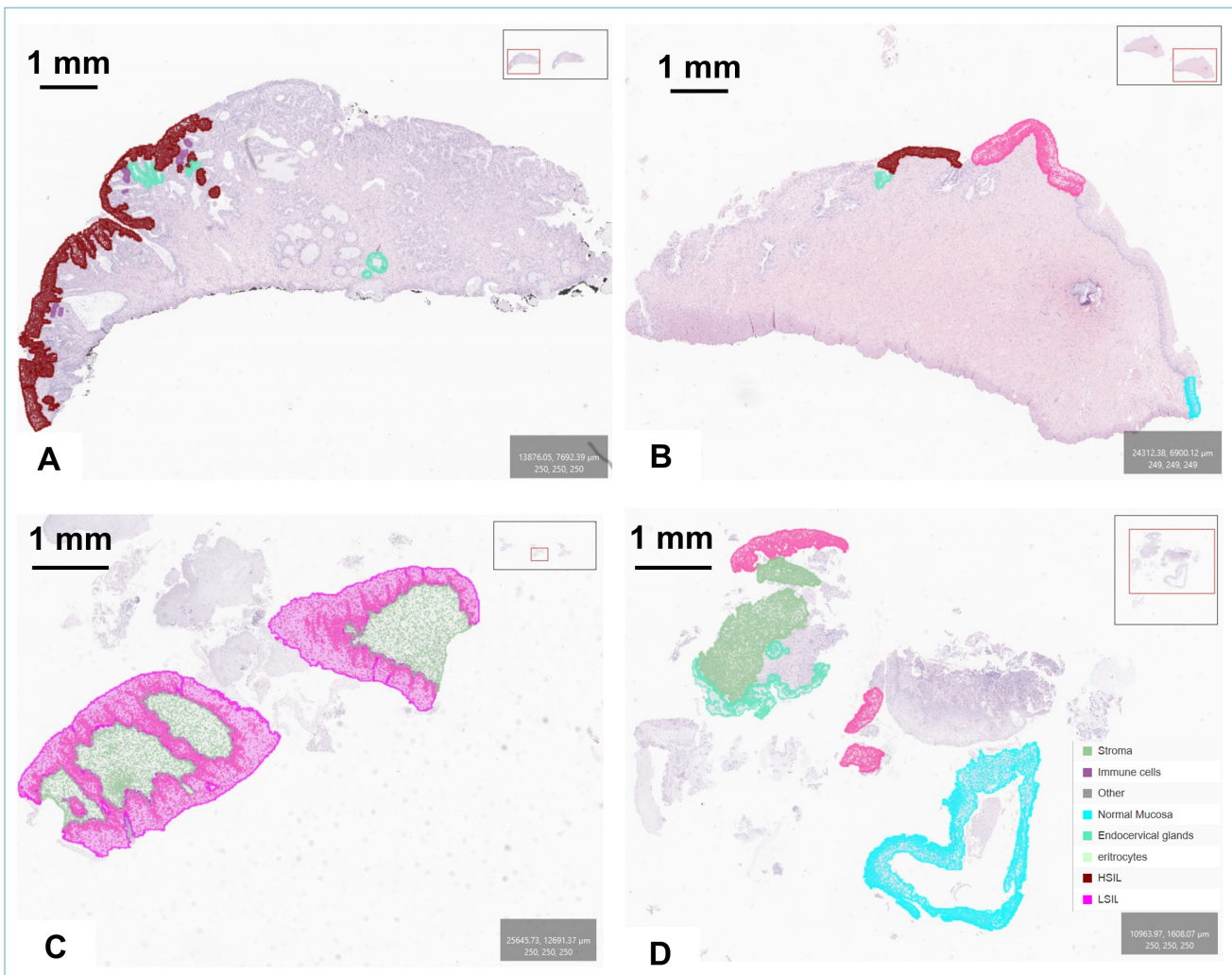


Figure 1. Examples of sparse pixel-level annotations used for training, illustrating the appearance and distribution of annotated regions across different cervical histology specimens. The images are shown in original histological staining with overlaid annotations of different colors for each tissue-type class. A) Example of an extensive HSIL involvement, few immune infiltrates and focal endocervical glands. B) Mixed LSIL/HSIL region with a little area of preserved normal mucosa. C) LSIL-dominant sections with clear stromal separation. D) Complex multi-region slide with four annotation classes represented.

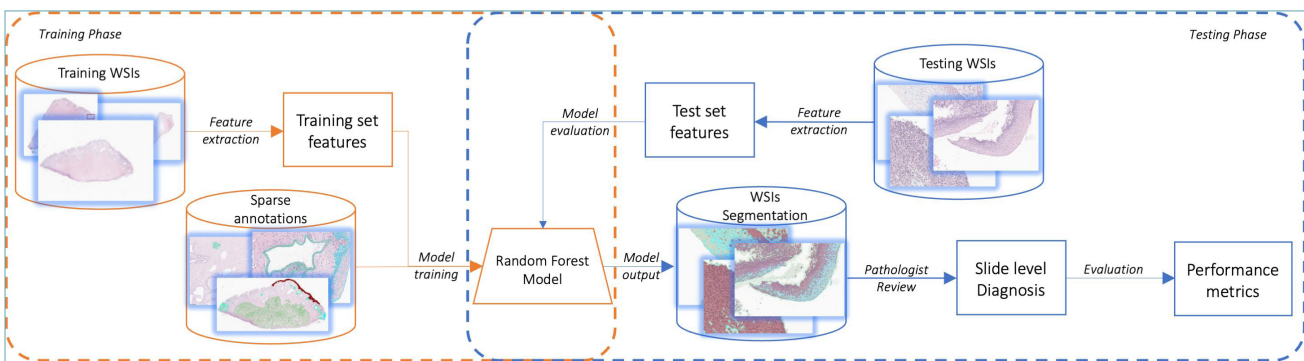


Figure 2. Schematic representation of the training and testing pipeline used in this study. In the training phase the features extracted from the training WSIs and the sparse annotations from the same WSIs are used to train the Random Forest classifier. The latter is used during the testing phase to produce a segmentation on the testing WSIs. Such segmentation is reviewed by the pathologist, who provides a slide-level diagnosis, which can be used to evaluate the performance of the semi-automated process.

overview of the computations workflow, distinguishing clearly the training phase and the testing phase, during which the segmentation maps are generated for expert review.

For the pixel-level classifier we selected Random Forest (RF) algorithm¹⁹ rather than DL or other ML classifiers. RFs offer several advantages that align with the characteristics of our data and the practical constraints of the intended deployment context. First, RFs scale efficiently to the millions of pixel-level samples generated from sparse annotations, without requiring extensive hyperparameter optimization or additional over/undersampling or data augmentation. Second, due to their bootstrap aggregation and feature subsampling, RFs exhibit inherent robustness to class imbalance and noisy annotations, which are common in epithelial tissue segmentation. Third, RF models are computationally lightweight, enabling both training and inference on standard CPU-only hardware, an essential requirement for low resource environments. Finally, RFs integrate natively within QuPath's classification framework, facilitating a reproducible and easily deployable workflow without additional software dependencies. To explore the trade-off between speed and level of detail, we developed two alternative variants of the pixel classifier that are referred to as Classifier 1 and Classifier 2. Both variants share the same feature extraction pipeline and serve the same goal, but differ in their classification granularity and intended use case.

The feature extraction step involves:

- Color deconvolution to isolate the three color channels (red, green and blue) and the stain components (e.g., hematoxylin and diaminobenzidine),
- Multiscale filters, including Gaussian smoothing, Laplacian, structure tensor coherence, and Hessian determinant,
- Feature normalization with standard scaling.

A total of 20 features per pixel were extracted. These features result from the combination of five image channels (Red, Green, Blue, and two color-deconvolved channels) with four multiscale filters (Gaussian, Laplacian, Structure Tensor Coherence, and Hessian Determinant). This compact feature set was chosen as it captures both color and structural information while remaining computationally efficient.

Classifier 1 was designed to perform a broad categorisation of tissue, distinguishing between three main classes: HSIL, LSIL and Normal epithelium. This model is optimized for rapid classification and offers coarse but clinically relevant segmentation of epithelial regions.

Classifier 2 extends the level of detail by segmenting tissue into eight categories: HSIL, LSIL, endocervical

mucosa, exocervical mucosa, immune cells, erythrocytes, stroma, and ignore. This detailed segmentation allows quantifying the spatial distribution of multiple tissue types and enables a more comprehensive assessment of the lesion.

Results

EXPERIMENTAL SETUP

Both above-described classifiers have been directly trained in a supervised fashion using QuPath's integrated OpenCV-based classifier framework. Specifically, we employed the RTrees classifier, an implementation of the RF algorithm provided by the OpenCV library, with the following hyperparameters:

- Decision trees number: 50,
- Maximum tree depth: 25,
- Minimum samples per leaf: 10.

Such hyperparameters were selected a priori rather than tuned through cross-validation. The selected parameters provided a good balance between model stability and computational efficiency, given the relatively limited number of features extracted per pixel and by the need to avoid tree fragmentation and overfitting, a common risk in decision tree models.

Each pixel is assigned one of the class labels used during training (3 or 8 for Classifier 1 and Classifier 2, respectively), each associated with a unique RGB color to facilitate downstream visualization and region quantification. Although the number of annotated WSIs was limited, each slide provides millions of training samples, making data augmentation unnecessary. This is consistent with prior work in similar pixel-wise classification tasks, where convergence can be achieved with a relatively small number of annotated pixels¹⁸. Even if at this stage we cannot release source code or annotated data, as the project is ongoing, the method can be reproduced in any QuPath installation using the parameters detailed above, since the pixel-level classifier relies exclusively on documented functionalities available to the community.

As previously mentioned, the segmentation maps provided by the pixel-level classifiers, were visually inspected by a pathologist, who assigned a slide-level diagnosis following the following criteria. Slides were evaluated as LSIL if they exhibited any of the following (Fig. 3, A-B):

- cervical epithelium classified as normal mucosa (turquoise), with only scattered pixels showing alternative classifications;
- all cervical epithelium classified as low-grade dysplasia (magenta);

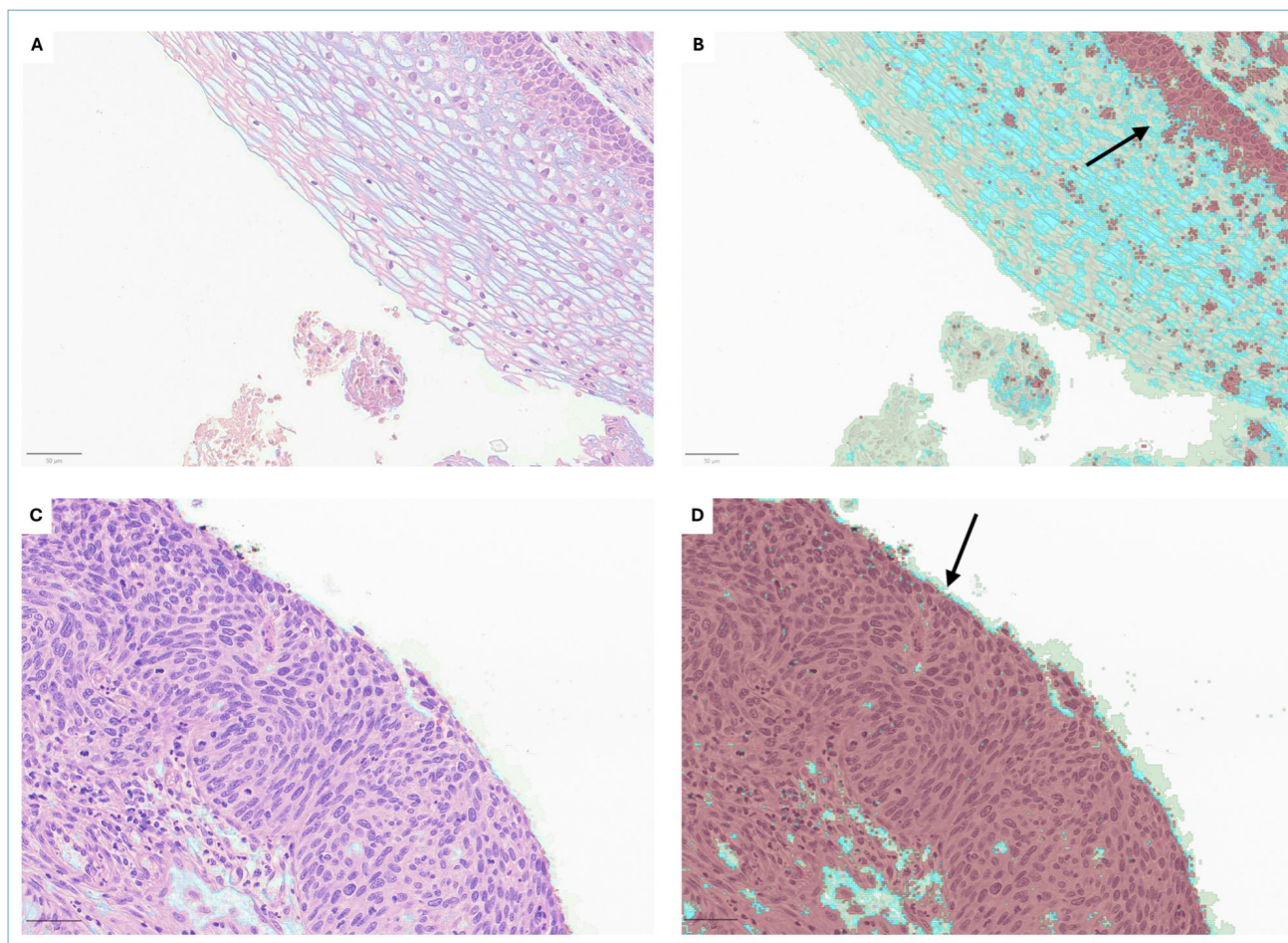


Figure 3. Examples of LSIL (A-B), and HSIL (C-D) are shown in original histological staining and corresponding classifier outputs (B and D); scale bar: 50 μ m. For clarity, only the output of Classifier 1 is shown. A-B) LSIL. The dysplastic epithelial layer is visible in the H&E image (C) and accurately identified by the classifier (D, arrow). C-D) HSIL. High-grade dysplasia involving the full epithelial thickness is visible in the original image (E) and confirmed by the classifier output (F, arrow).

- cervical epithelium classified as high-grade dysplasia (bordeaux) limited to the lower third of the epithelial thickness;
- any combination of the above within the same slide.

Slides were evaluated as HSIL if they exhibited, cervical epithelium at any point classified as high-grade dysplasia (bordeaux) involving more than one-third of the epithelial thickness (Fig. 3, C-D).

The obtained diagnoses were then compared with the reference annotations to compute performance metrics. Since this is a binary classification task, we defined HSIL as the positive (P) class and LSIL as the negative (N) class. Accordingly:

- True Positive (TP): an HSIL case correctly classified as HSIL;
- False Positive (FP): an LSIL case incorrectly clas-

sified as HSIL;

- True Negative (TN): an LSIL case correctly classified as LSIL;
- False Negative (FN): an HSIL case incorrectly classified as LSIL.

From a clinical perspective, our primary objective is to minimize the number of FN, as missing a high-grade lesion could lead to delayed diagnosis and treatment. For this reason, in addition to the classic accuracy, we included precision, recall, and F1-score in our evaluation, as these metrics provide a more comprehensive assessment of the model's diagnostic reliability, especially in imbalanced or high-risk scenarios.

EXPERIMENTAL RESULTS

On the 309 test set slides, both classifiers achieved remarkable performance; Classifier 1 produced 10 FP

Table II. Performances obtained with the proposed classifiers in the diagnosis of LSIL and HSIL.

Method	Accuracy	Precision	Recall	F1-score
Classifier 1	0.9644	0.9359	0.9932	0.9636
Classifier 2	0.9741	0.9653	0.9949	0.9799

and Classifier 2 produced 7 FP, with both models yielding a single FN. The overall key metrics are reported in Table II, and as can be observed, the classifiers reached 96.44% and 97.41% accuracy, respectively. While both classifiers exhibit strong classification capabilities, Classifier 2 slightly outperforms Classifier 1

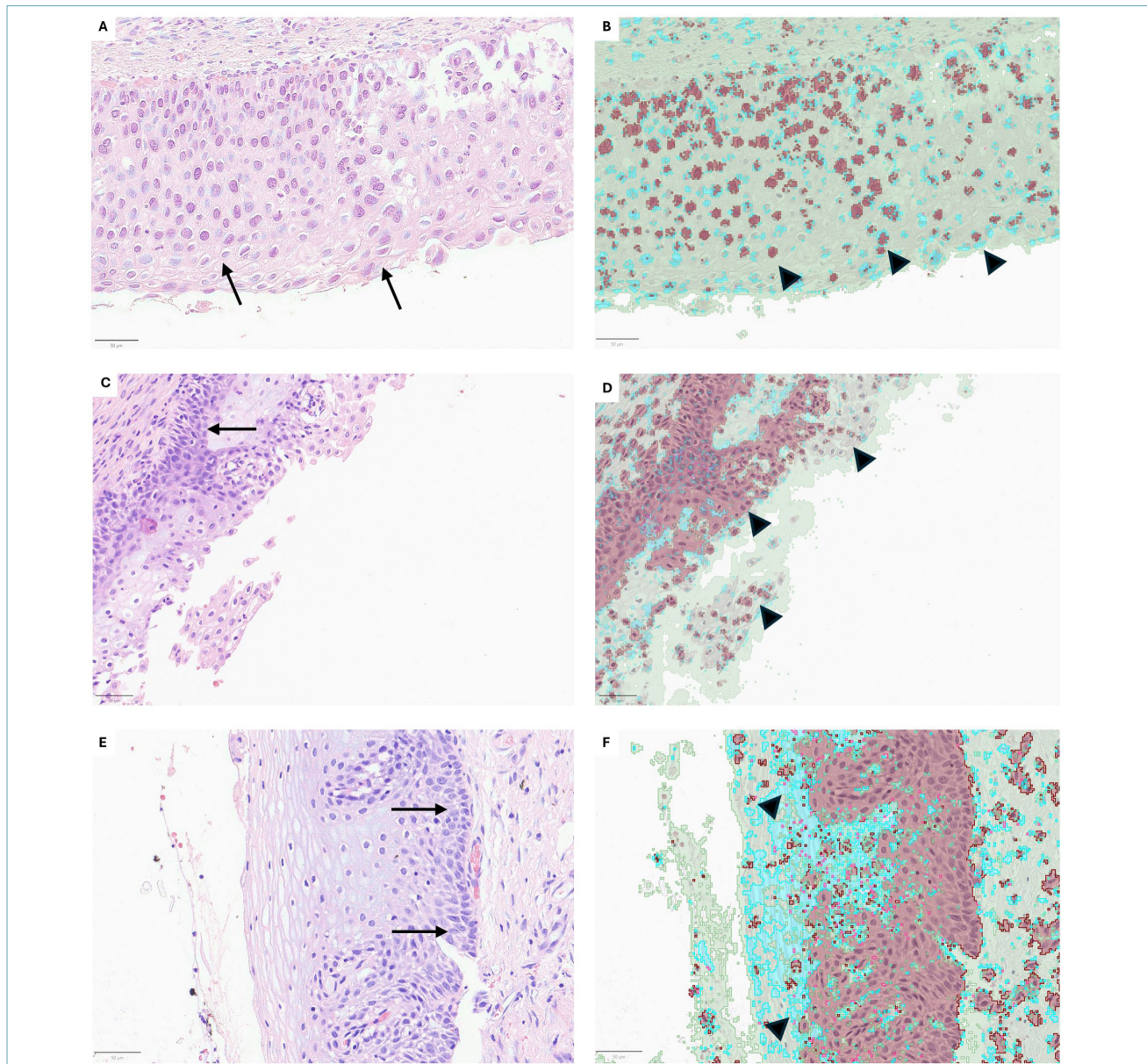


Figure 4. Examples of failure cases: FN (A-B), FP (C-D), and FP (E-F) shown in original histological staining (A, C and E) and corresponding classifier outputs (B, D and F). Scale bar: 50 μ m. For clarity, only the output of Classifier 1 is shown. A-B) FN. In A, H&E staining shows high-grade dysplasia extending through the full thickness of the mucosa. In B, the same region is misclassified by the software as LSIL. Although scattered high-grade dysplastic foci are present even in the upper layers (arrowheads), the lack of compactness and homogeneity led the classifier to underestimate the severity. C-D) FP. In C, H&E staining reveals a compact dysplastic area restricted to the lower epithelial layers (arrows), consistent with LSIL, as later confirmed by p16 immunohistochemistry (not shown). In D, the same region is erroneously classified as HSIL by the software, which misinterpreted the extent of dysplasia up to the superficial layers (arrowheads). E-F) FP. In E, LSIL with basal dysplasia (arrows) is shown. In F, epithelial papillae (arrowheads) cause an apparent elevation of the epithelium, resulting in a distorted architecture. This papillomatosis led the classifier to misinterpret the lesion as high-grade dysplasia involving more than two-thirds of the epithelial thickness (arrowheads), thus overestimating its severity.

Table III. Comparison among existing methods focusing on cervical lesions using histological images and the proposed classifiers. * Study comparable to ours.

Reference	Dataset/ Task	Accuracy	Precision	Recall	F1-score
Li C (2019)	Private (Not specified)	0.88	0.8796	0.8787	0.8786
Jiang (2022)	Private (CIN grades)	0.8241	-	0.8333	0.8241
Song* (2022)	Private (SCC/AC)	0.917	-	0.912	0.932
Li Y (2023)	Private (SCC/AC)	0.96	0.969	0.976	0.944
Our (Classifier 1)	Private (LSIL/HSIL)	0.9644	0.9359	0.9932	0.9636
Our (Classifier 2)	Private (LSIL/HSIL)	0.9741	0.9653	0.9949	0.9799

in all key metrics. Its superior precision and recall indicate it is more dependable for detecting TP while minimising both FN and FP, making it the preferred model in this comparison. Representative failure cases are shown in Figure 4, including the single FN and two FP. Comparing our results with those reported in previous studies on histological image analysis proves inherently difficult for several reasons. Firstly, all prior works have relied on proprietary datasets, preventing direct benchmarking. Moreover, the classification tasks addressed differ substantially from ours; most studies focused on distinguishing between distinct carcinoma types, such as SCC and AC, rather than grading intraepithelial lesions. The only study that presents a task somewhat comparable to ours – indicated in the table with an asterisk – relied on the outdated CIN-based classification system, further limiting the validity of a direct comparison. Nonetheless, we attempted a quantitative comparison and reported in Table III a selection of representative methods that focused on cervical lesions using histological images, alongside the same performance metrics previously described. As shown in the table, both classifiers we propose demonstrate notably high performance compared to existing approaches, particularly in terms of recall, underscoring their potential in minimizing FN in a diagnostic setting.

Discussion

TECHNICAL FEATURES

At all stages of the process, from processing to digitisation, an identical protocol was sought for all samples. In the choice of the cutting thickness of the paraffin blocks, the best compromise was 4 microns, as 3 microns were not sufficiently colored. Beyond 4 microns, it was possible to obtain a greater depth of the image, with intense coloring, but the digitized file was too heavy. During the scan, it was realized that there is a need to find the ideal slide format specifically created for the scanner, as well as for image analysis. One of the biggest technical problems was the slowness in reading the images. Despite being equipped with excellent hardware and software, the use of histological images was particularly cumbersome. Moreover, the use of a single scanner and uniform staining workflow, while ensuring excellent internal consistency, reduces data variability and must be considered when interpreting the experimental results.

Emergent technical limits of the study

The technical limitations of the present study revolve around the need for a change in habits and organisation within pathological anatomy services. First of all, we were not able to take advantage of the resolute staff, so the same protocol was conducted by technicians who deal with other sectors almost all of their time. Despite the use of automated stainers and scanners, these were not connected in a single process chain and were not built to work in the same workflow. The slides were assembled manually, as we did not have a fitter available. This logically highlighted an inhomogeneity of the pressure applied to the slide, resin, and powder residues and a variable orientation. These conditions then led to some defects in the scanned image: cut edges of the image, difficulty in reading any further sections set in addition to the central one, and slight changes in focus that could have disturbed the analysis of the pixels. In addition, the characteristics of the computer used for the analysis of the image were not sufficient to obtain a speed of analysis suitable for a large department of pathological anatomy. However, all the above technical and organizational constraints reflect real world conditions typical of low resource laboratories, but inherently introduce heterogeneity in slide preparation and scanning quality, which may affect reproducibility.

Beyond these technical and organizational aspects, further challenges must be considered for a potential clinical implementation of the method. Regulatory approval remains essential before integration into routine workflows, requiring compliance with international standards for medical software and data protection. Successful clinical adoption also depends on user

trust: although the pixel-level output is inherently interpretable and aligned with the diagnostic reasoning of pathologists, adequate training and gradual workflow integration will be necessary. Finally, infrastructural variability across institutions – including scanner types, staining protocols, and computing resources – may challenge widespread adoption, especially in resource-constrained settings, which are precisely the environments where low-cost solutions are most needed.

Emerged metodological limits of the study

Beyond technical considerations, several methodological limitations must also be acknowledged. The entire dataset was collected at a single institution and processed using a uniform workflow, scanner, and staining protocol. While this homogeneity ensured internal consistency, it inevitably restricts the variability of the material available for training and testing, thus limiting the generalizability of the results to settings that differ in laboratory procedures or acquisition devices.

Because the WSIs were fully deidentified at source, no patient-level metadata were available, and it was therefore not possible to enforce a patient-level split or to distinguish between specimen types such as biopsies and conizations. The train–test partition had to be performed at the slide level, with the inherent possibility that slides originating from the same (unknown) patient might appear in both sets. However, given the limited number of patients (60 in total), enforcing a patient-based split would not necessarily have improved robustness: in fact, it could have reduced inter-patient

variability even further and increased the risk of overfitting to patient-specific patterns. Moreover, although specimen type information could not be recovered, the unavoidable mixing of biopsies and conizations introduces a degree of natural variability in epithelial morphology, which, while not explicitly controllable, may slightly mitigate overfitting by exposing the model to a broader range of tissue presentations. As larger and coded datasets become available, a more stringent patient-level allocation will be feasible.

Another important limitation concerns the diversity rather than the quantity of training data. Although the training set contained only 42 WSIs, each slide contributed millions of annotated pixels, providing a substantial amount of training signal for the Random Forest classifier. The constraint lies not in the size of the pixel-level data but in its relatively narrow representativeness, due to both the single center design and the retrospective nature of the cohort.

Current advances and the value of weak supervision

The adoption of AI in pathology is accelerating. Advances such as virtual staining, automation, and image-based molecular inference require a shift from a microscope-centered workflow to a multidisciplinary model involving pathologists, morphologists, engineers, computer scientists, robotics experts, and data managers (Fig. 5). AI based tools can also support training by providing consistent annotations, rare case simulations, and interactive feedback, particularly valuable in settings with limited teaching resources.

AI has shown the potential to approach or even sur-

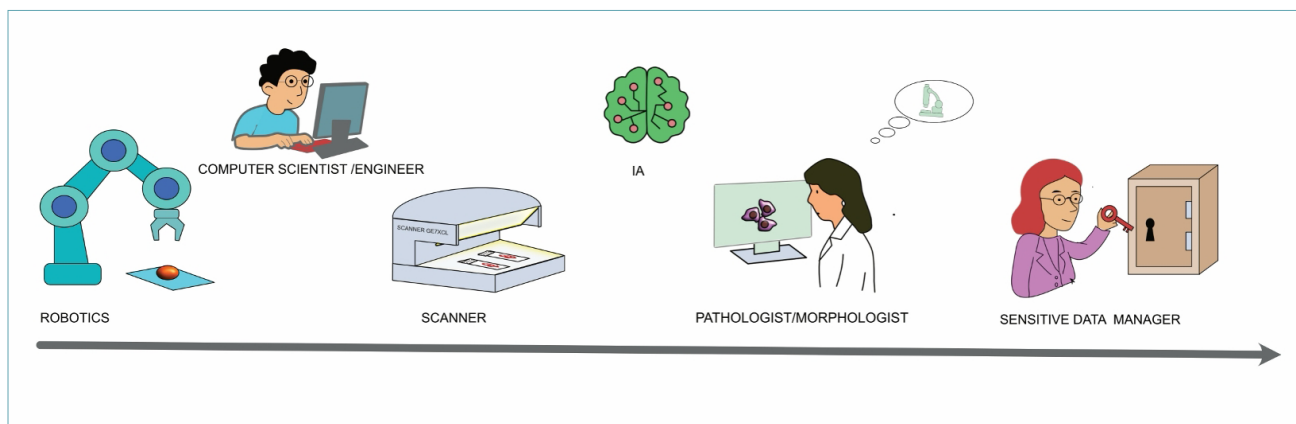


Figure 5. Workflow of a multidisciplinary team of modern and future pathology. Some figures, such as the computer scientist and the sensitive data manager, are still rare but will be fundamental in the future. Robotics is only in its infancy and will likely be comparable to what has already been introduced in clinical pathology automation. It is likely that the figure of the pathologist will have to be strengthened in a morphological sense and integrated more and more with other figures of morphologists (clinical anatomists) to offer their skills in the conception of annotations and the increasingly fine interpretations of structures necessary for the instruction of the software.

pass pathologists in specific diagnostic tasks^{22–24}. Nevertheless, the development of robust AI models requires large-scale annotated datasets, which are costly and time-consuming to produce. Our results demonstrate that sparse pixel-level annotations, even from a limited number of WSIs, can support accurate and interpretable segmentation.

On the independent test set of 309 slides concordance reached 96.44% and 97.41%. The FN rate (only 1 case) is particularly encouraging, as it highlights the method's ability to reliably detect clinically significant lesions while maintaining a low FP rate; 10 and 7 for Classifier 1 and Classifier 2, respectively.

COMPARISON WITH STATE-OF-THE-ART METHODS

Recent advances in histopathological AI have been dominated by DL architectures, including CNNs, transformers, and attention-based models. These methods offer powerful feature extraction capabilities and are particularly effective in complex tasks such as pixel-wise segmentation and epithelial boundary detection. However, their deployment requires extensive computational resources, large annotated datasets, and specialized expertise – factors that can hinder their adoption, especially in low-resource healthcare environments. In contrast, our RF-based approach shows that lightweight and explainable models can still achieve competitive performance, with the added benefit of accessibility. By leveraging open-source tools such as QuPath and OpenCV, our framework prioritises scalability and healthcare equity, ensuring that AI-assisted diagnostic tools can be realistically implemented across diverse clinical settings.

FUTURE PROSPECTS

Our results offer a very encouraging scenario. Through the creation of pixel and cell-detection algorithms built on QuPath, it will be possible to offer a low-cost tool that can potentially be applied to all departments, including those with scarce resources available. Looking ahead, the expansion of the dataset to multiple centers and acquisition workflows will be essential to evaluate the robustness and generalizability of the proposed approach across heterogeneous staining, scanning, and laboratory conditions. As additional annotated material becomes available, the pixel-level segmentation pipeline can be progressively refined and incorporated into routine diagnostic and quality assurance workflows, supporting pathologists through interpretable and reproducible tissue maps. Furthermore, the structured pixel-level outputs may represent a valuable didactic resource, enabling targeted visualization of diagnostic patterns and supporting training activities without modifying the methodological scope

of the present study.

Conclusions

In conclusion, the results support the use of open-source software as a diagnostic and didactic support tool, capable of increasing efficiency and standardisation in slide reading. However, the role of the pathologist in the final validation remains fundamental today, especially in borderline or ambiguous cases and where there is a lack of infrastructure, dedicated technical personnel and equipment that do not allow optimal standardization of the digital process.

CONFLICTS OF INTEREST STATEMENT

The authors declare no conflicts of interest.

FUNDING

This research received no external funding.

AUTHOR CONTRIBUTIONS

Conceptualization, Laura Nonnis, Marcello Trucas, Clara Gerosa; Data curation, Laura Nonnis, Stefano Guerriero, Marco Palomba, Alessio Rocca; Formal analysis, Laura Nonnis; Investigation, Laura Nonnis, Michela Vincis, Daniela Fanni; Methodology, Laura Nonnis and Lorenzo Putzu; Software, Laura Nonnis and Lorenzo Putzu; Supervision, Lorenzo Putzu, Clara Gerosa and Marcello Trucas; Validation, Laura Nonnis, Lorenzo Putzu and Michela Vincis; Visualization, Laura Nonnis and Marcello Trucas; Writing – original draft, Lorenzo Putzu and Marcello Trucas; Writing – review & editing, Laura Nonnis, Lorenzo Putzu, Michela Vincis, Daniela Fanni, Clara Gerosa and Marcello Trucas.

ETHICAL CONSIDERATION

The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was waived since the analysis involved only anonymised retrospective data from patients who had given consent at the time of surgery.

References

- 1 Nicolás I, Saco A, Barnadas E, et al. Prognostic implications of genotyping and p16 immunostaining in HPV-positive tumors of the uterine cervix. *Modern Pathology* 2020;33:128-137; <https://doi.org/10.1038/s41379-019-0360-3>
- 2 Nayar R, Wilbur DC. The Bethesda System for Reporting Cervical Cytology: A Historical Perspective. *Acta Cytol* 2017;61:359-372. <https://doi.org/10.1159/000477556>
- 3 Trucas M, Vincis M, Intini C, et al. The potential translational utility of embalmed cadaveric gastrointestinal tract specimens:

- a proof-of-concept study. *Translational Research in Anatomy* 2025;39:100404. <https://doi.org/10.1016/j.tria.2025.100404>
- 4 Dolciotti C, Righi M, Greco E, et al. The translational power of Alzheimer's-based organoid models in personalized medicine: an integrated biological and digital approach embodying patient clinical history. *Front Cell Neurosci* 2025;19:1553642. <https://doi.org/10.3389/fncel.2025.1553642>
 - 5 Trucas M, Dervan A, Quondamatteo F. Are spinal cord and medulla samples from embalmed donors suitable for histological examination? A pilot study. *J Anat* 2024;246:1042-1052. <https://doi.org/10.1111/joa.14188>
 - 6 Littera R, Perra A, Miglianti M, et al. The double-sided of human leukocyte antigen-G molecules in type 1 autoimmune hepatitis. *Front Immunol* 2022;13: 1007647. <https://doi.org/10.3389/fimmu.2022.1007647>
 - 7 Berbis MA, McClintock DS, Bychkov A, et al. Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. *EBioMedicine* 2023;88:104427. <https://doi.org/10.1016/j.ebiom.2022.104427>
 - 8 Jantzen J, Norup J, Dounias G, et al. Pap-Smear Benchmark Data For Pattern Classification. In: *Proc. NiSIS 2005 : Nature Inspired Smart Information Systems (NiSIS), EU Co-Ordination Action DTU orbit; 2005.*
 - 9 Pacal I. MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection. *Knowl Based Syst* 2024;289:111482. <https://doi.org/10.1016/j.knosys.2024.111482>
 - 10 Hussain E, Mahanta LB, Borah H, et al. Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. *Data Brief* 2020;30:105589. <https://doi.org/10.1016/j.dib.2020.105589>
 - 11 Rezende MT, Silva R, Bernardo F de O, et al. Cric searchable image database as a public platform for conventional pap smear cytology data. *Sci Data* 2021;8:151. <https://doi.org/10.1038/s41597-021-00933-8>
 - 12 Yang X, Li C, He R, et al. CAISHI: A benchmark histopathological H&E image dataset for cervical adenocarcinoma in situ identification, retrieval and few-shot learning evaluation. *Data Brief* 2024;53:110141. <https://doi.org/10.1016/j.dib.2024.110141>
 - 13 Li C, Chen H, Zhang L, et al. Cervical Histopathology Image Classification Using Multilayer Hidden Conditional Random Fields and Weakly Supervised Learning. *IEEE Access* 2019;7:90378-90397. <https://doi.org/10.1109/ACCESS.2019.2924467>
 - 14 Jiang P, Liu J, Wang L, et al. Classifying Cervical Histopathological Whole Slide Images via Deep Multi-Instance Transfer Learning. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) IEEE; 2022;2302-2307; https://doi.org/10.1109/BIBM55620.2022.9995014*
 - 15 Song J, Im S, Lee SH, et al. Deep Learning-Based Classification of Uterine Cervical and Endometrial Cancer Subtypes from Whole-Slide Histopathology Images. *Diagnostics* 2022;12:2623. <https://doi.org/10.3390/diagnostics12112623>
 - 16 Li Y, Chen F, Shi J, et al. Convolutional Neural Networks for Classifying Cervical Cancer Types Using Histological Images. *J Digit Imaging* 2022;36:441-449. <https://doi.org/10.1007/s10278-022-00722-8>
 - 17 Alquran H, Azani Mustafa W, Abu Qasmieh I, et al. Cervical Cancer Classification Using Combined Machine Learning and Deep Learning Approach. *Computers, Materials & Continua* 2022;72:5117-5134. <https://doi.org/10.32604/cmc.2022.025692>
 - 18 Di Ruberto C, Loddo A, Putzu L. A leucocytes count system from blood smear images. *Mach Vis Appl* 2016;27:1151-1160. <https://doi.org/10.1007/s00138-016-0812-4>
 - 19 Breiman L. Random forests. *Mach Learn* 2001;45:5-32. <https://doi.org/10.1023/A:1010933404324/METRICS>
 - 20 Kulkarni PA, Singh H. Artificial Intelligence in Clinical Diagnosis. *JAMA* 2023;330:317. <https://doi.org/10.1001/jama.2023.11440>
 - 21 Beam AL, Drazen JM, Kohane IS, et al. Artificial Intelligence in Medicine. *New England Journal of Medicine* 2023;388:1220-1221. <https://doi.org/10.1056/NEJMe2206291>
 - 22 Campanella G, Kumar N, Nanda S, et al. Real-world deployment of a fine-tuned pathology foundation model for lung cancer biomarker detection. *Nat Med* 2025;31:3002-3010. <https://doi.org/10.1038/s41591-025-03780-x>
 - 23 Huang Z, Yang E, Shen J, et al. A pathologist-AI collaboration framework for enhancing diagnostic accuracies and efficiencies. *Nat Biomed Eng* 2024;9:455-470. <https://doi.org/10.1038/s41551-024-01223-5>
 - 24 Verma KK, Palmal R, Yadav P. Unveiling the future: the impact of artificial intelligence in diagnostic pathology. *Surgical and Experimental Pathology* 2025;8:23. <https://doi.org/10.1186/s42047-025-00197-1>