



Automatic screening system for conventional Pap smear analysis

David Kupas^a , Ilona Kovacs^b, Andras Hajdu^a , Balazs Harangi^a

^a Department of Data Science and Visualization, Faculty of Informatics, University of Debrecen, Kassai ut 26, Debrecen, 4028, Hajdu-Bihar, Hungary

^b Department of Pathology, University of Debrecen Clinical Center, Kenézy Gyula Campus, Bartok Bela ut 2-26, Debrecen, 4031, Hajdu-Bihar, Hungary

ARTICLE INFO

Keywords:

Cervical cancer screening
Deep learning
Ensemble classification
Pap smear analysis

ABSTRACT

This paper presents a comprehensive automatic screening system designed to assist cytological experts in the evaluation of Pap smears for cervical cancer detection. This work presents an end-to-end automatic screening system developed specifically for conventional Pap smears. The goal of the system is to reduce the burden on human experts by providing a reliable, resource-efficient, and accurate solution that can distinguish between normal and diseased smears. The proposed automatic screening system processes whole-slide images by first dividing them into 2000×2000 pixel tiles, then applies a YOLO-based localization algorithm to extract individual cells. These cells are subsequently classified by an ensemble of nine deep neural networks, including convolutional and transformer-based architectures, trained on the publicly available APACC dataset. The final stage involves aggregating cell-level predictions across models to generate smear-level features, which are used to train a LightGBM classifier. The system is evaluated on a private dataset of 339 expert-annotated smears collected in a clinical setting. Despite the relatively small number of smears, over 12 million individual cell predictions were utilized, resulting in an average smear-level classification accuracy of 88.8% using cross-validation. These findings demonstrate the feasibility and potential of such automated systems to support existing cervical cancer screening workflows and to enhance diagnostic consistency and throughput.

1. Introduction

Cervical cancer remains a leading cause of cancer-related mortality in women globally, particularly in low- and middle-income countries [1]. According to the World Health Organization, effective screening and early treatment can significantly reduce its incidence and mortality [2]. For decades, the Papanicolaou (Pap) smear test has been the cornerstone of screening programs, offering a simple yet effective method for detecting precancerous cellular changes [3].

However, conventional Pap smear screening is a labor-intensive process that depends heavily on expert cytotechnologists and pathologists. Manual examination of whole-slide images (WSIs) is time-consuming and prone to inter- and intra-observer variability, which can cause diagnostic inconsistencies [4,5]. Concurrently, the number of trained cytopathologists is declining due to regulatory, training, and economic pressures, making manual evaluation increasingly unsustainable.

This situation underscores an urgent need for automated systems to ensure consistent, high-quality screening with minimal human intervention. The advent of digital pathology has spurred interest in using artificial intelligence (AI) to automate Pap smear analysis [6]. Deep learning, particularly convolutional neural networks (CNNs), has shown great promise in medical imaging, including both cytological

and histopathological analysis [7]. More recently, transformer-based architectures have been introduced, offering improved performance by modeling long-range dependencies and global contextual features [8].

Earlier approaches often relied on handcrafted features and classical machine learning, but these have been largely supplanted by end-to-end deep learning systems [7,9]. Several studies have proposed CNN-based architectures for classifying individual cervical cells. For instance, Zhang et al. introduced a cascaded CNN model for Pap smear classification that achieved promising results on benchmark datasets [9]. Others have developed systems for Liquid-Based Cytology (LBC) WSIs, demonstrating effective slide-level classification in large cohorts [10].

Beyond individual models, some researchers have explored WSI-level analysis using patch-based methods and hybrid architectures combining CNNs and transformers [11]. However, most prior work has focused on individual image or patch classification. Far fewer studies have addressed the challenge of aggregating cell-level predictions for a whole-smear diagnosis. Simple approaches like majority voting or averaging may not capture the complex spatial or probabilistic patterns present in the data.

* Corresponding author.

E-mail addresses: kupas.david@inf.unideb.hu (D. Kupas), kovacs.ilona@med.unideb.hu (I. Kovacs), andras.hajdu@inf.unideb.hu (A. Hajdu), harangi.balazs@inf.unideb.hu (B. Harangi).

<https://doi.org/10.1016/j.bspc.2026.110716>

Received 5 November 2025; Received in revised form 13 May 2026; Accepted 25 May 2026

1746-8094/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

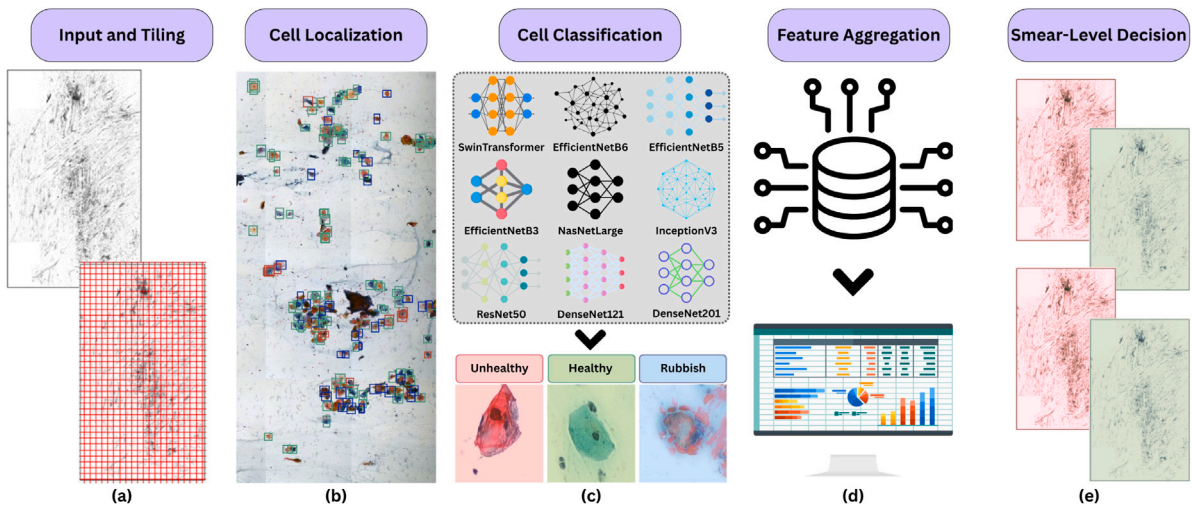


Fig. 1. Overview of the end-to-end pipeline for Pap smear-based cervical cancer screening. (a) Raw smear digitization, and smear tiling into fixed-size patches. (b) Cell localization via YOLOv10x. (c) Classification of each cell using a nine-model ensemble of CNNs and transformers. (d) Threshold-based aggregation of model outputs into engineered features. (e) Final smear-level prediction using a LightGBM classifier.

In recent years, several AI-powered systems have emerged to support or replace cytological examination, most notably those based on LBC. High-profile commercial systems, such as Hologic's Genius Digital Diagnostics System, have shown strong performance in clinical evaluations [12].

Despite recent progress, a key constraint persists: systems deployed in practice are built primarily around LBC, limiting applicability in settings where conventional Pap smears remain the standard due to cost and infrastructure considerations. As a result, LBC-specific solutions are often impractical in these environments.

Unlike conventional smears, LBC involves specimen processing that reduces obscuring material and enriches diagnostically useful cells; in contrast, this study addresses conventional smears where debris, blood, and cellular overlap must be managed by the pipeline. As such, we propose a comprehensive, fully automated deep learning pipeline tailored to conventional Pap smear analysis. A high-level overview of the pipeline is shown in Fig. 1. Our pipeline is designed to operate end-to-end from WSI processing and cell localization to cell classification and final smear-level diagnosis. The system builds on publicly available datasets and is evaluated with real-world, clinically validated data. The final architecture includes an ensemble of convolutional and transformer-based neural networks and leverages aggregation techniques and feature selection to enhance reliability.

The rest of this paper is structured as follows. Section 2 introduces the datasets used for training and evaluation. Section 3 details the methodology, including the localization strategy, cell classification, and smear-level decision making. Section 4 presents the experimental results, while Section 5 discusses the clinical implications, limitations, and future directions of our approach.

2. Datasets

The development of the proposed system relies on three primary datasets, each serving a specific role within the pipeline. These include public sources that collectively support model development from cell-level localization to whole-smear classification.

2.1. Cell detection dataset

To support the training and validation of the cell localization component, we utilized three publicly available datasets: CRIC [13], CCEDD [14], and a custom-processed subset of APACS [15]. All three

datasets consist of cervical cytology images stained using the Papanicolaou method. The APACS dataset is explicitly described as conventional Pap smear WSIs. The CRIC dataset is also indicated as derived from conventional Pap smear cytology data. For CCEDD, liquid-based cytology preparations were used, the exact cytology platform is not specified. These datasets contain high-resolution cropped fields, providing a rich foundation for training robust detection algorithms.

The CRIC Cervix dataset consists of 400 cropped image fields derived from conventional Pap smears using Zeiss AxioImager Z2 (40× objective, 10× eyepiece, AxioCam MRC), annotated with 11,534 individual cell masks. Each image includes multiple annotated cells according to the Bethesda system, supporting multi-class segmentation tasks. The dataset is widely used for instance-level training and includes detailed metadata for each case.

The CCEDD dataset includes 686 high-resolution images collected from patient samples stained using the Papanicolaou method, scanned with a Nikon ELIPSE Ci slide scanner, SmartV350D lens and a 3-megapixel digital camera. For patients with negative and positive cervical cancer, the optical magnification is 100× and 400×, respectively. A total of 33,614 cells were annotated with detailed boundaries into two classes (normal or abnormal). The annotations were performed by six expert cytologists and include segmentation masks for each cell. This dataset serves as a high-quality source for evaluating detection algorithms under realistic staining and morphological conditions.

The APACS dataset comprises Pap smear slices extracted from 5 full-slide WSIs of conventional Pap smears scanned with a 3DHISTECH Panoramic 1000 scanner at 200× magnification. The dataset includes expert pixel-wise annotations of all visible cells. For the purpose of this study, we employed a watershed-based segmentation pipeline to convert pixel-level masks into bounding boxes. This process enabled the construction of a large annotated training corpus with high diversity in cell density, smear quality, and staining artifacts.

A detailed breakdown of the annotated training dataset composition is shown in Table 1, including the number of smear slices and bounding boxes per dataset and split.

2.2. Cell classification dataset

For cell-level classification, we employed the APACC dataset [16], which consists of 103,675 manually annotated cervical cell images extracted from 107 digitized conventional Pap smear WSIs. The WSIs were scanned using a 3DHISTECH Panoramic 1000 scanner with a 20× objective lens (achieving the 200× magnification level as in

Table 1

Distribution of annotated smear slices and bounding boxes across datasets and splits.

Dataset	Split	Smear slices	Bounding boxes
CCEDD	Training	552	28,097
	Validation	64	3153
	Test	70	3072
CRIC	Training	319	9258
	Validation	42	1263
	Test	39	1013
APACS	Training	2324	29,646
	Validation	256	3510
	Test	286	3779
Total	-	3,952	82,791

Table 2

Training–test split of the APACC dataset used for cell classification.

Class	Training samples	Test samples	Total
Healthy	28,895	5826	34,721
Unhealthy	2366	576	2942
Rubbish	50,371	11,757	62,128
Total	81,632	18,159	99,791

the APACS dataset), producing MRXS-format images. These images were divided into 2000×2000 pixel tiles, and cell candidates were automatically segmented using a fully convolutional neural network pipeline.

Each extracted cell was independently annotated by three expert cytopathologists into four classes: healthy (NILM: normal cytology), unhealthy (Epithelial cell abnormality: ASC/LSIL/HSIL, no sub-division), rubbish (Unsatisfactory for evaluation: non-informative debris/artifacts, overlapping/folded cells, blood/leukocytes, bacteria, or obscuring material), and both cells (mixed healthy/unhealthy). Rubbish dominates due to exhaustive automatic extraction without filtering. For training of the classification models, only three classes were retained: healthy, unhealthy, and rubbish. A detailed breakdown of the APACC dataset split is shown in Table 2. The dataset is hosted on the Open Science Framework (OSF) and is publicly available under a permissive license [17].

Although the APACC dataset was used for classifier training, it is already composed of extracted individual cell images rather than full-slide patches. Therefore, the classifier is trained on cell-level crops, which makes the training representation closer to the deployment setting than a slide-level or mask-level formulation. The remaining difference lies in how the crops are generated: APACC crops are derived from automatic extraction, whereas deployment crops are produced by the YOLO-based localization module.

2.3. Smear-level evaluation dataset

For the validation of the system, we collected a new dataset consisting of 339 digitized conventional Pap smears. The smears were acquired and scanned using the same protocol as the APACC dataset. Each smear corresponds to an individual patient and was annotated by expert cytopathologists. Normal/diseased classification was based on expert visual cytological assessment. All abnormal (positive) cases were further confirmed by histology. The normal (negative) cases were derived from pre-screening. This dataset serves as the clinical reference for evaluating whole-smear decision making.

3. Methodology

We propose a fully automated system that processes Pap smear WSIs to produce smear-level evaluation regarding their normal or diseased status. The pipeline begins by reading MRXS files, efficiently handling

the large images in memory. Each WSI is divided into tiles of size 2000×2000 pixels to create manageable regions for downstream analysis. These tiles are then individually processed by a localization model, which detects and localizes individual cervical cells within each tile. For every detected bounding box, a cropped image of the cell is extracted directly from the corresponding tile.

The extracted cell images are passed through a deep learning ensemble classifier with each member model producing predicted class probabilities for all input cells, classifying them as healthy, unhealthy, or rubbish (unusable). In contrast to traditional ensemble averaging, our approach preserves all individual model outputs to support a more nuanced feature construction process in later stages.

In the final phase, smear-level features are engineered from the aggregated model predictions. These features are derived from the predicted probability distributions generated by the models for individual cells and incorporate statistics computed across different class thresholds. The resulting feature vectors are then used as input for a simple, lightweight classifier, which is trained to make binary predictions at the smear level. This entire pipeline, from raw whole-slide input to final classification, operates without any manual intervention.

3.1. Cell detection stage

The purpose of this stage is to detect and extract individual cervical cells from high-resolution smear tiles (2000×2000 pixels), which serve as the visual basis for subsequent classification. Instead of relying on pixel-level segmentation masks during inference that is often computationally intensive and less scalable, we opted for a cell localization strategy using bounding boxes. This allowed us to adopt a fast and cost-efficient object detection solution while maintaining cell-level granularity.

To train the localization model, we constructed a large composite dataset combining annotated cell data from three public sources: CRIC [13], CCEDD [14], and a custom subset of the APACS dataset [15]. Both the CRIC and CCEDD datasets contain explicitly annotated cell regions suitable for detection model training. However, the APACS dataset originally provides only binary segmentation masks for each full-slide image tile.

To adapt the APACS dataset for our detection pipeline, we implemented a custom watershed-based method that converts dense binary masks into individual bounding boxes. This included preprocessing steps to filter out artifacts, peak detection for separating overlapping regions, and watershed segmentation guided by a distance transform. Peak detection uses a multimodal approach on the distance transform (local maxima seeds) to robustly separate cell clusters/overlaps, while initial artifact filtering applies unimodal area thresholding, that has been chosen empirically for realism without over-segmentation. This procedure enabled the reuse of high-quality segmentation data in a more efficient detection framework. The detailed steps of this transformation are described in Algorithm 1.

To visually support the description of our method, we include an example input and intermediate outputs of the conversion process in Fig. 2. This demonstrates the peak detection and watershed segmentation steps that lead to bounding box extraction.

The resulting bounding boxes from all datasets were unified into a single training corpus, which served as the input for training our cell localization model. This approach allowed us to leverage multiple sources while avoiding inconsistencies between pixel-level and box-level annotations.

During inference, the classifier receives cell crops generated by the YOLO localization module. Thus, both training and deployment operate on cropped cell images, although the crop source differs. This design reduces, but does not completely eliminate, the possibility of a domain shift between annotated training crops and detector-derived crops.

Algorithm 1 Watershed-Based Cell Extraction for the APACS Dataset

Input: Binary segmentation mask for each smear tile
Output: Filtered bounding boxes for object detection

```

1: for each connected region in the segmentation mask do
2:   Compute area and convex area
3:   if area < 2000 or area / convex_area ≥ 0.95 then
4:     Discard region
5:   end if
6: end for
7: Compute the distance transform of the binary mask
8: Detect local maxima in the distance map:
   • Use 3 × 3 footprint
   • Minimum distance between peaks = 30 pixels
   • Allow max 2 peaks per region
9: Apply watershed segmentation using:
   • Inverted distance map as topography
   • Local maxima as seed points
10: for each resulting segment do
11:   Compute bounding box coordinates
12:   Convert to YOLO-compatible format
13: end for
14: Save cropped cell images and annotation files

```

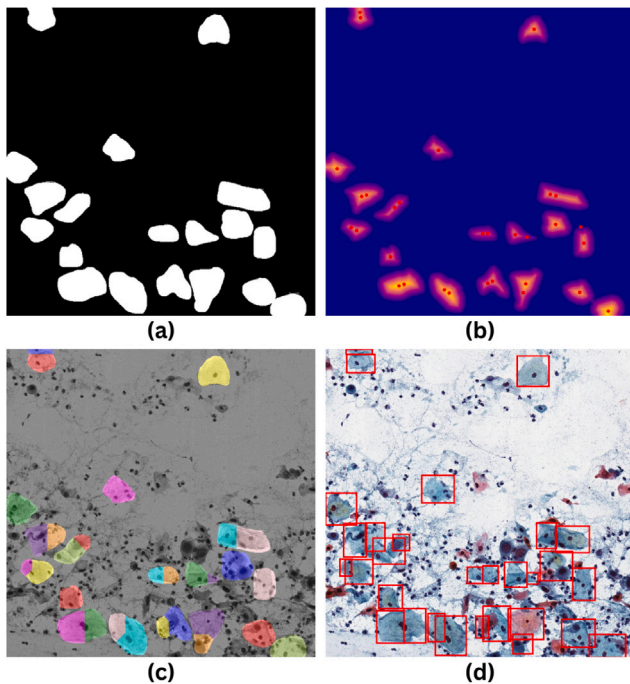


Fig. 2. Intermediate stages of watershed-based cell extraction. (a) Original binary mask in the APACS dataset. (b) Distance transform with local maxima (red dots) used for seed generation. (c) Watershed segmentation overlaid on the original image. (d) Final YOLO-compatible bounding boxes.

3.2. Cell classification stage

This stage is responsible for classifying each localized cell as either healthy, unhealthy, or irrelevant (rubbish). To achieve robust and diverse predictions, an ensemble of nine independently trained deep learning models was utilized. The ensemble includes eight convolutional neural networks (CNNs), as well as one vision transformer:

DenseNet121 and DenseNet201. [18] The DenseNet family introduces dense connectivity between layers, which improves gradient flow and feature reuse. DenseNet121 provides a compact, efficient architecture widely used in medical imaging tasks. DenseNet201, a deeper variant, offers enhanced feature extraction capacity, which is useful for capturing subtle differences in cervical cell morphology.

EfficientNetB3, B5, and B6 [19]. EfficientNet models scale network width, depth, and resolution using a compound scaling method. We selected B3, B5, and B6 to explore the trade-off between accuracy and computational complexity across different scales. These models are known for their strong performance on image classification benchmarks and are particularly suited for medical tasks, where image features are fine-grained and class imbalance is common.

ResNet50 [20]. ResNet50 is a well-established CNN architecture featuring residual connections, which mitigate the vanishing gradient problem and facilitate training deeper models. Its robustness and moderate size make it a practical baseline for comparison in biomedical tasks.

InceptionV3 [21]. InceptionV3 employs multi-scale convolutional blocks that help capture features at different spatial resolutions. This is especially beneficial in cytological classification, where cell morphology varies significantly across samples.

NASNetLarge [22]. NASNetLarge is a neural architecture search-derived model, optimized on large-scale classification tasks. Its automated design yields a highly expressive model with state-of-the-art performance, though at higher computational cost. Including NASNet in the ensemble introduces a top-performing, structurally diverse model.

Swin transformer [23]. The Swin Transformer adapts transformer-based architectures to vision tasks by introducing hierarchical feature maps and shifted windows. Transformers are well-suited to medical imaging because of their capacity to model global context. Swin was fine-tuned from an ImageNet-pretrained checkpoint using our APACS dataset, offering a complementary representation to CNN-based models.

Each model processes cropped cell images generated during the detection stage. Before inference, a standardized aspect-ratio-preserving preprocessing is applied: cell images are padded with white pixels (value 255) to form square inputs and resized to 224×224 pixels. This ensures morphological integrity during resizing. Standard ImageNet normalization is applied to all inputs, with preprocessing tailored to each architecture via its official implementation. All models output a softmax-normalized class probability distribution over the three categories. These outputs are stored without any ensemble voting or averaging, preserving their diversity for use in the downstream smear-level classification module.

3.3. Smear-level decision

Our initial attempt at smear-level classification relied on the ratio between the numbers of unhealthy and healthy cells per smear, excluding cells predicted as rubbish. This approach did not incorporate model confidence scores or engineered features.

Fig. 3 illustrates the distributions of unhealthy cell ratios for four representative models. These models were selected due to their diverse architectural properties and representative behavior; the rest of the ensemble showed similar patterns, albeit on slightly different scales.

Each model is represented by two boxplots (candle plots): green for smears annotated as normal, and red for those classified as diseased. The plotted values correspond to the proportion of cells classified as unhealthy relative to healthy (note that cells predicted as rubbish were excluded from this calculation).

In a perfectly separable scenario, red boxplots (diseased smears) would consistently show higher unhealthy ratios than their green counterparts. While the tendency of diseased smears generally exhibiting

Table 3
Smear-level AUC for simple pooling strategies.

Strategy	Healthy AUC (%)	Unhealthy AUC (%)
max	46.6	61.0
mean	47.5	51.8
median	46.7	48.8
std	36.6	54.8
frac>0.5	45.3	54.5
sum	39.2	40.9

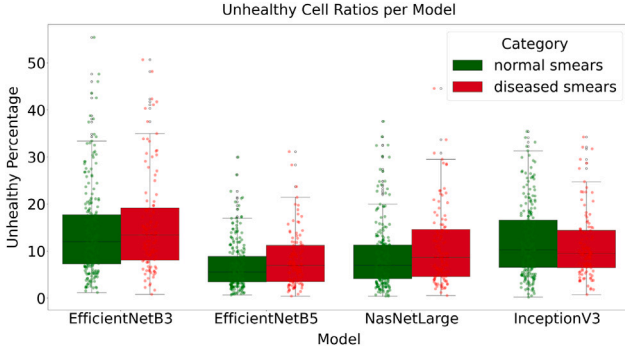


Fig. 3. Distributions of predicted unhealthy cell ratios for four representative models across smears. Each model has two boxplots: green for normal smears and red for diseased smears. Values indicate the percentage of unhealthy cells (excluding rubbish) relative to healthy cells within a smear.

higher predicted unhealthy percentages is observable, it is also notable, that the overlap between the two distributions remains substantial for all four models. A simple threshold-based decision using these ratios serves as a natural baseline. Using stratified 10-fold CV with per-fold F1-optimized thresholds on training data only, it achieved accuracy $35.4\% \pm 4.6\%$, F1 0.490 ± 0.028 , AUC 0.543 ± 0.103 (mean threshold: 0.030). These poor results confirm substantial distribution overlap and demonstrate the value of our threshold-based feature construction, which captures model-specific confidence patterns.

Simple pooling of cell-level probabilities shows limited discriminatory power. Table 3 reports AUC for common aggregation strategies applied to unhealthy/healthy probabilities across all 9 models per smear (max AUC=61.0%). These substantially underperform our threshold-count features, confirming their contribution.

To obtain a more discriminative representation of smear-level information, we employed a threshold-based feature extraction strategy. Each cell is associated with nine softmax probability vectors, one from each of the nine classification models. Rather than averaging these vectors, we retain them independently to enable richer, model-specific feature engineering.

Smear-level features are generated by aggregating model predictions across all cells within a smear. For each model, class, and confidence threshold, count-based features are computed indicating how many cells surpass the threshold for that class. Thresholds are uniformly sampled from 0.05 to 0.95 in steps of 0.05, and the focus is placed on the diagnostic classes (healthy and unhealthy). This uniform grid ensures model-agnostic feature construction that is robust to calibration differences across models. Repeating this for each ensemble member yields features that capture how frequently cells are confidently assigned to each diagnostic class by each model.

Formally, the procedure can be described as follows. Let $s = \{c_1, c_2, \dots, c_n\}$ denote the set of detected cells in a smear, and let $\mathbf{p}_i^{(m)} = [p_{i,\text{healthy}}^{(m)}, p_{i,\text{rubbish}}^{(m)}, p_{i,\text{unhealthy}}^{(m)}]$ represent the softmax probabilities output by model m for cell c_i . Let $\mathcal{T} = \{0.05, 0.10, \dots, 0.95\}$ denote the set of thresholds and $\mathcal{C} = \{\text{healthy}, \text{unhealthy}\}$ the classes of interest. For each model m , class $c \in \mathcal{C}$, and threshold $t \in \mathcal{T}$, the engineered feature

is defined as

$$f_{c,t}^{(m)}(s) = \sum_{i=1}^n \mathbb{I}(p_{i,c}^{(m)} > t),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function (which equals to 1 if the condition holds and 0 otherwise). This construct yields a high-dimensional and interpretable feature vector that characterizes nuanced patterns in model predictions across the ensemble. These features are subsequently employed as input to a downstream classifier responsible for smear-level diagnostic decisions.

These feature vectors were used as input to a LightGBM classifier [24], trained to distinguish between normal and diseased smears. The training process incorporated automated hyperparameter optimization and feature selection using the Optuna framework [25].

Feature selection was performed internally using LightGBM's built-in feature importance evaluation. Specifically, we ranked features based on their contribution to the model's objective function, meaning the total gain attributed to each feature across all decision tree splits. This gain-based importance measure reflects how much each feature contributes to reducing the loss function during training.

During the optimization phase, we evaluated multiple configurations using only the top-ranked features and selected the final model based on cross-validation performance and generalization ability.

All evaluation results are detailed in Section 4, where we also list the selected features used in the final configuration.

4. Experimental results

This section evaluates the proposed pipeline at both component and system levels. We begin by analyzing the performance of the individual modules, namely the YOLO-based cell localization and the image-level classification ensemble. Subsequently, we present the final smear-level classification results, supported by both quantitative metrics and relevant visualizations.

4.1. Component-level evaluation

4.1.1. Cell localization

The YOLOv10x architecture builds upon previous iterations of the YOLO family, offering enhanced real-time detection capabilities through an improved backbone and optimized training procedures [26]. The model was trained from scratch using a composite dataset consisting of CRIC, CCEDD, and APACS smear images, with the latter preprocessed using a custom watershed-based cell extraction method (as described in Section 3.1). The aim of this stage was to detect and extract individual cervical cells from high-resolution smear tiles to support downstream classification.

Standard object detection metrics such as mean Average Precision (mAP) were used to quantify model performance during training, validation, and testing. During training, the loss values decreased steadily as well as validation precision, recall, and mAP increasing accordingly, confirming consistent convergence without evident overfitting. The best validation results of 77% precision, 80% recall and 85% mAP at IoU 50% were reached at the 128th epoch.

On the independent test set, the model achieved a precision of 76%, recall of 79%, F1 score of 78%, and a fitness score of 65%. The mAP was 85% at IoU 50%, 72% at IoU 75%, and 63% when averaged over IoU thresholds from 50% to 95%.

These results demonstrate robust detection performance across a range of localization thresholds. Beyond these quantitative metrics, our ultimate evaluation emphasized an application-driven criterion: ensuring that the model could reliably detect sufficient numbers of cells in complete smears for accurate downstream analysis.

Qualitative visual inspection confirmed high coverage across smears with varying cell density and staining conditions. To further assess the robustness of this stage, we conducted an ablation-style experiment

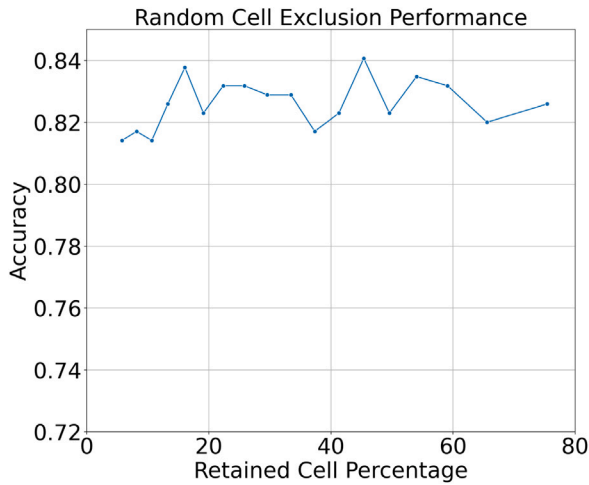


Fig. 4. Smear-level classification accuracy as a function of the proportion of retained cells.

by reducing the number of available cell images prior to classification through random exclusion. Specifically, we evaluated the effect of randomly discarding varying proportions of detected cells on overall smear-level classification accuracy using a simple non-optimized random forest classifier.

As shown in Fig. 4, the system maintained stable smear-level classification accuracy even under high levels of random filtering, indicating resilience to partial loss of cell detections.

These findings suggest that predictive information is broadly distributed throughout the smear, and that the cell localization module provides a robust and redundant input for downstream analysis.

4.1.2. Cell classification

Each of the nine models used in our evaluation was trained independently using well-established convolutional and transformer-based architectures. All networks were initialized with weights pretrained on ImageNet, allowing them to benefit from general-purpose visual features. For each model, the top classification layers were replaced with a custom head consisting of a global average pooling layer, followed by a dense layer with 512 units and ReLU activation, a dropout layer with a rate of 50%, and a final dense layer with softmax activation outputting class probabilities.

Input images were processed in a standardized way prior to training. Each image was first padded with white background pixels as necessary to preserve the original aspect ratio, then resized to 224×224 pixels. This approach ensured that the spatial relationships and morphological features within each image were retained without distortion. Image normalization was applied using preprocessing routines specific to each model family.

During training, extensive data augmentation was employed to promote generalization and robustness. Augmentation techniques included random horizontal and vertical flipping, random rotation, zooming, and color variations. All augmentations were designed to be biologically plausible and consistent with the expected variability in cell appearance.

Optimization for all models was performed using either the Adam optimizer, with architecture-specific learning rate schedules and batch sizes adjusted according to available GPU memory. Early stopping was applied based on validation performance to prevent overfitting. Each model was trained and evaluated separately on the predefined training and test splits. Evaluation metrics included accuracy, macro F1-score, and area under the receiver operating characteristic curve (AUC), and were computed using predictions on the test set.

Table 4

Individual model performance on the APACC test set.

Model	Accuracy (%)	F1 score (%)	AUC (%)
SwinTransformer	85.25	85.26	86.02
EfficientNetB6	84.28	83.14	84.40
EfficientNetB5	83.79	84.02	84.13
EfficientNetB3	82.12	83.43	85.08
NasNetLarge	81.85	82.64	83.46
InceptionV3	81.41	82.73	84.42
ResNet50	81.39	80.75	81.71
DenseNet121	80.03	79.07	80.31
DenseNet201	79.87	78.92	80.25

Table 5

Performance of various models trained on the full feature set without parameter optimization.

Model	Accuracy (%)	F1 score (%)	AUC (%)
MLP	85.84	85.51	85.29
Random Forest	84.66	83.89	86.33
XGBoost	84.96	84.47	88.29
LightGBM	87.61	87.19	88.17
Stacking	86.73	86.39	84.96
Voting	84.96	84.47	88.27

Despite differences in architecture and parameter count, most models achieved closely comparable performance across all metrics as presented in Table 4.

4.2. Smear-level evaluation

This section evaluates the final stage of the pipeline, where aggregated predictions from cell-level classifiers are used to determine the diagnostic result of each smear. We report baseline comparisons, model selection experiments, and final classification results.

4.2.1. Classifier comparison

To identify the most suitable model for smear-level classification, we compared several commonly used classifiers on a consistent set of engineered features derived from cell-level predictions. These features capture the number of cells per smear for which a specific model predicted a given class (healthy or unhealthy) above a defined confidence threshold, as described in Section 3.3.

The evaluated classifiers included Multilayer Perceptron (MLP), Random Forest, XGBoost, LightGBM, and ensemble-based methods such as voting and stacking. The MLP used in the comparison is a custom deep neural network implemented in PyTorch, consisting of four fully connected layers with 256, 128, 64, and 1 neuron per layer, respectively, together with batch normalization, ReLU activation, and dropout at rates of 0.3, 0.2, and 0.2 between the first three hidden layers. Tree-based models (Random Forest, XGBoost, and LightGBM) and the ensemble variants (voting and stacking) were evaluated using default hyperparameters.

As shown in Table 5, LightGBM achieved the best overall performance, with an accuracy of 87.61%, an F1 score of 87.19%, and an AUC of 88.17%. While stacking achieved a similar F1 score, its slightly lower AUC and higher computational cost made LightGBM a more practical choice. In summary, the LightGBM-based smear-level ensemble delivers the strongest overall performance; this supports the expectation that aggregating model-specific confidence patterns across many cells provides additional discriminatory power beyond any single cell classifier. Based on these results, all subsequent evaluations were conducted using LightGBM.

Table 6

Top 17 features selected for the final model based on LightGBM internal gain. Each feature represents the number of cells for which the given model predicted the specified class with a probability greater than the listed threshold.

Model	Predicted class	Lower bound Threshold (%)
DenseNet201	healthy	85
ResNet50	healthy	90
DenseNet201	unhealthy	15
DenseNet121	unhealthy	30
ResNet50	healthy	85
NASNetLarge	healthy	95
DenseNet201	healthy	80
DenseNet201	healthy	90
DenseNet121	healthy	85
DenseNet121	healthy	90
Swin Transformer	unhealthy	95
InceptionV3	healthy	85
DenseNet201	healthy	75
Swin Transformer	unhealthy	5
InceptionV3	unhealthy	90
DenseNet121	unhealthy	95
DenseNet201	unhealthy	20

4.2.2. Feature-based classification and final results

As described in Section 3.3, a structured feature representation was derived by aggregating the predicted class probabilities across thousands of detected cell images. While preliminary experiments suggested that tree-based ensemble models perform better than simple classifiers (e.g., SVM, logistic regression, or MLP) on this feature set, LightGBM emerged as the most promising choice in terms of both performance and efficiency.

To reduce redundancy and improve generalizability, we applied internal gain-based feature importance ranking using the LightGBM framework. Based on this ranking, the top 17 most informative features were selected for the final model based on empirical evaluation. Multiple feature set sizes were tested and it could be observed, that the performance plateaued around 15–20 features, with 17 providing the best performance. These features are binary count-based representations, each indicating how many cells in a smear were predicted with high confidence (above a certain threshold) for a specific class by a specific model.

Table 6 lists these top-ranked features, organized by model, predicted class, and applied threshold.

We evaluated the optimized LightGBM model using stratified 10-fold cross-validation on the full set of 339 manually annotated smears (abnormal prevalence: 31.9%). The model achieved an average accuracy of $88.8\% \pm 7.3\%$, F1 score of $88.1\% \pm 8.1\%$, AUC of $89.2\% \pm 9.3\%$, with sensitivity $96.6\% \pm 2.3\%$ (95% CI), specificity $72.2\% \pm 12.3\%$ (95% CI), PPV $88.8\% \pm 4.7\%$ (95% CI), and NPV $90.3\% \pm 7.4\%$ (95% CI). These confidence intervals were derived from per-fold metrics using the standard normal approximation ($z = 1.96 \times \sigma / \sqrt{10}$). Notably, PPV and NPV are prevalence-dependent; our cohort's elevated prevalence (31.9%) yields higher PPV than typical screening populations, where PPV would decrease while NPV would increase further, supporting safe rule-out decisions.

In this cross-validation approach, the dataset was divided into ten stratified folds, ensuring that each subset maintained the original class distributions. The model was trained and tested ten times, each time holding out a different fold as the validation set and using the remaining nine folds for training. The reported metrics represent the mean and standard deviation calculated across these ten runs. This provides a robust estimation of the model's ability to generalize and illustrates the performance variance arising from different data splits.

To assess the classifier's decision boundaries, Fig. 5 shows the confusion matrix aggregated over all folds, while the ROC curve produced from cross-fold predictions is presented in Fig. 6.

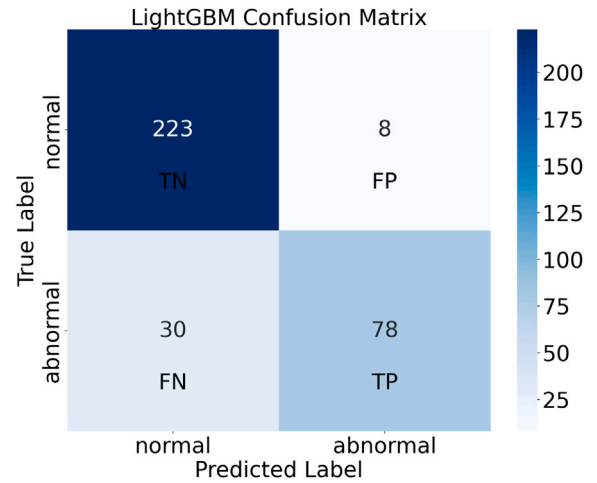


Fig. 5. Confusion matrix for the best-performing smear-level model (LightGBM, 17 features), aggregated over all cross-validation folds.

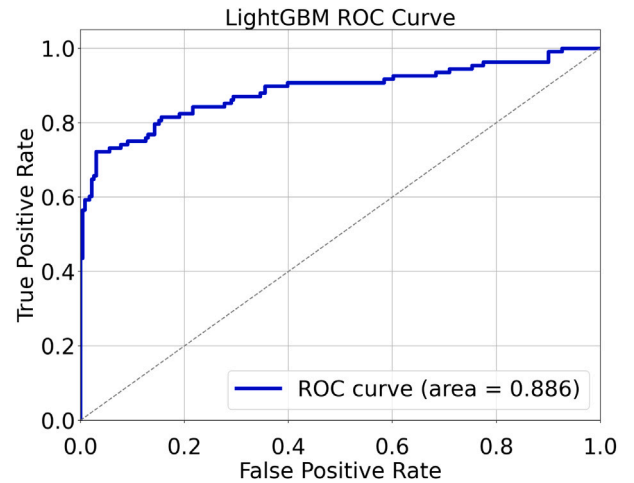


Fig. 6. ROC curve for smear-level classification using the optimized LightGBM model, using predictions from cross-validated folds.

5. Discussion

5.1. Performance and clinical relevance

The presented automatic screening system demonstrates strong clinical potential, achieving an average smear-level classification accuracy of $88.8\% \pm 7.3\%$, F1 score of $88.1\% \pm 8.1\%$, and AUC of $89.2\% \pm 9.3\%$ using stratified 10-fold cross-validation on 339 expert-annotated clinical smears (31.9% diseased prevalence). The high sensitivity of $96.6\% \pm 2.3\%$ makes the system particularly suitable for screening applications where minimizing false negatives is paramount, while the specificity of $72.2\% \pm 12.3\%$ maintains reasonable workload for expert review of positive cases.

A key strength lies in the threshold-based feature aggregation, which substantially outperformed simpler baselines. Simple unhealthy/healthy ratios yielded only 35.4% accuracy and 54.3% AUC, while common pooling strategies (max/mean/median) peaked at 61.0% AUC. The final LightGBM model using the top-17 threshold-count features achieved a performance plateau, demonstrating that model-specific confidence patterns across the 9-model ensemble provide rich discriminatory power beyond basic statistics.

The system's robustness to imperfect cell detection indicates that diagnostic signal is broadly distributed across smears rather than concentrated in sparse regions maintaining smear accuracy even when randomly discarding up to 50% of detections. This redundancy supports reliable performance even under challenging imaging conditions or partial occlusions typical of conventional Pap smears.

5.2. Component-level analysis

Cell localization using the YOLOv10x achieved strong detection metrics (mAP@0.5=85%, precision 76%, recall 79%, F1 78%) on the composite dataset. The custom watershed preprocessing of APACS pixel masks successfully converted dense segmentations into YOLO-compatible boxes while preserving cluster/overlapping cell diversity. Multimodal peak detection on distance transforms proved robust for cytology-specific clustering patterns.

Cell classification implemented by the 9-model ensemble (8 CNNs + 1 Swin Transformer) delivered consistent cell-level performance (80%–85% F1, 81%–86% AUC on APACC test set), with architectural diversity reducing correlated errors. EfficientNetB3 emerged as optimally balanced for accuracy versus computation difficulty. Standardized preprocessing (white padding to 224×224 , ImageNet normalization) and biologically plausible augmentations ensured morphological fidelity across models.

Smear-level aggregation using the uniform 0.05-step threshold grid (0.05–0.95) across healthy/unhealthy classes for each model created model-agnostic features robust to calibration differences, enabling LightGBM to exploit ensemble complementarity effectively.

5.3. Limitations

Despite strong performance, several limitations warrant consideration. First, a residual domain shift exists between APACC training crops (FCN-segmented) and deployment crops from YOLO localization, potentially affecting crop tightness, background content, or partial cell inclusion. While cell-level F1 scores remain high (80%–85%), direct quantification on a manually annotated YOLO crop subset would clarify this effect.

Second, the validation cohort ($n=339$ smears) is relatively small with elevated disease prevalence (31.9%), boosting PPV/NPV relative to typical screening populations. Multi-center validation across diverse demographics, staining variations, and scanners remains essential for generalizability.

Third, the “rubbish” class dominates APACC due to exhaustive unfiltered extraction, which is realistic for conventional smears but requires robust ensemble handling of debris/blood/overlaps. While excluded from smear features, extreme imbalance could still influence cell classifiers.

Fourth, smear-level baselines focus on classical aggregation (pooling/ratios) rather than recent WSI deep-aggregation methods, which target LBC/histopathology WSIs with cleaner data characteristics. Direct re-implementation on our private cohort risks unfair comparison absent shared datasets; our threshold aggregation's superiority over 6 pooling strategies (AUC 89.2% vs max 61.0%) validates the approach for conventional smear challenges.

5.4. Computational complexity and deployment readiness

The end-to-end pipeline processes WSIs fully automatically, averaging 12 min per smear on dual NVIDIA A100 80 GB GPUs + AMD EPYC 7742 64-core CPU (research implementation). Breakdown reveals ensemble inference as the primary bottleneck, followed by tiling/YOLO localization, with feature engineering/LightGBM negligible.

This conservative runtime excludes production optimizations like TensorRT acceleration, INT8 quantization, model pruning, or ONNX

export, suggesting potential 3–5× speedup. The modular design supports parallel tile processing and model batching, while lightweight LightGBM enables CPU deployment. It is also worth noting, that the system used for research implementation is also ready for training the described systems. If the inference is the only task, a high-performing PC can also perform the necessary computations. Early tests show, that a high-end consumer PC with maximized memory is also plausible, although with increased processing time.

5.5. Comparison to prior work

Unlike LBC-focused systems, the described conventional Pap smear pipeline addresses cost-constrained settings where LBC remains impractical. While cell-level classifiers match literature, the key advance lies in smear-level aggregation: threshold features outperform majority voting/pooling, approaching clinical benchmarks on realistic unfiltered data.

The 9-model ensemble exceeds single-model approaches, with architectural diversity (CNN+transformer) providing complementary error patterns absent in uniform architectures. YOLOv10x localization (85% mAP@0.5) surpasses traditional segmentation for deployment scalability.

5.6. Future directions

Immediate priorities include multi-center validation (target $n > 2000$ smears) spanning demographics, scanners, and staining variations to support regulatory approval. Domain adaptation techniques could mitigate training-deployment crop differences.

Algorithmic enhancements encompass spatial aggregation (cell density maps, cluster statistics), self-supervised pretraining on unlabeled WSIs, and uncertainty quantification for triage. Production deployment requires CE marking pathway validation.

Long-term, expanding to glandular lesions and HPV+ stratification could evolve the system into a comprehensive cervical screening platform, supporting global elimination targets through accessible automation.

6. Conclusions

This study delivers a fully automated end-to-end screening system for conventional Pap smears, achieving 88.8% CV accuracy, 89.2% AUC, and 88.1% F1 on 339 clinically annotated WSIs—leveraging YOLOv10x localization (mAP@0.5=85%), 9-model CNN+transformer ensemble (80%–85% cell F1), and LightGBM on threshold-confidence features from 12M+ predictions.

Key innovations include: (1) watershed adaptation of APACS pixel masks for robust localization training, (2) model-diverse ensemble aggregation outperforming baselines, (3) deployment-feasible 12min/smear runtime. The system maintains high sensitivity (96.6%) suitable for screening while excluding realistic “rubbish” debris from smear decisions.

Results validate deep learning's role in supporting cytologists, particularly for resource-constrained conventional Pap workflows versus LBC-centric commercial systems. Multi-center validation will enable regulatory approval and clinical integration, with the potential to advance equitable cervical cancer screening globally.

CRedit authorship contribution statement

David Kupas: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Ilona Kovacs:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Andras Hajdu:** Writing – review &

editing, Validation, Supervision, Software, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Balazs Harangi:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethical approval

The research conducted at the University of Debrecen, Debrecen, Hungary, received approval from the Scientific and Research Ethics Committee of the Health Sciences Council of Hungary, referred to later as IRB, under protocol number OGYÉI/65989/2020. The data collected from anonymous samples preclude the identification of patients. Consequently, participants were not obligated to provide consent for data sharing. The IRB has waived and delegated the authority to publish and approve this work to the project leaders, namely Prof. Dr. Andras Hajdu and Dr. Ilona Kovacs. In accordance with the agreements outlined above, publication approval necessitates the endorsement of at least one of these project leaders, adhering to the ethical standards set forth in the World Medical Association's Helsinki Declaration and the University of Debrecen's scientific application regulations for ethical requirements in scientific publications.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Perplexity and DeepL in order to improve sentence formulation, clarity, and grammatical correctness throughout the paper. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the project TKP2021-NKTA-34 with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund of Hungary under the TKP2021-NKTA funding scheme. The authors would like to thank the cytopathologists Zoltán Hargitai and Zita Szombathy for their valuable contributions to the annotation process and clinical validation as well as all fellow colleagues who contributed to the project in any way. The authors also acknowledge the support of the Department of Pathology, University of Debrecen Clinical Center, Kenézy Gyula Campus, as well as the Faculty of Informatics at the University of Debrecen, for providing the infrastructure and resources necessary to carry out this work.

Data availability

The data that has been used is confidential.

References

- [1] Paul A. Cohen, Anjua Jhingran, Ana Oaknin, Lynette Denny, Cervical cancer, *Lancet* 393 (10167) (2019) 169–182.
- [2] World Health Organization, WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention, 2021, ISBN: 9789240030824. URL <https://www.who.int/publications/i/item/9789240030824>.
- [3] Rosemary H. Tambouret, The evolution of the papanicolaou smear, *Clin. Obstet. Gynecol.* 56 (1) (2013) 3–9.
- [4] Joann G Elmore, Gary M Longton, Patricia A Carney, Berta M Geller, Tracy Onega, Anna NA Tosteson, Heidi D Nelson, Margaret S Pepe, Kimberly H Allison, Stuart J Schnitt, et al., Diagnostic concordance among pathologists interpreting breast biopsy specimens, *Jama* 313 (11) (2015) 1122–1132.
- [5] Ivana Kholová, Giovanni Negri, Maria Nasioutziki, Laura Ventura, Arrigo Capitanio, Massimo Bongiovanni, Paul A Cross, Claire Bourgain, Henrik Edvardsson, Rosario Granados, et al., Inter-and intraobserver agreement in whole-slide digital ThinPrep samples of low-grade squamous lesions of the cervix uteri with known high-risk HPV status: a multicentric international study, *Cancer Cytopathol.* 130 (12) (2022) 939–948.
- [6] Michael Matheny, Sonoo Thadaneey Israni, Mahnoor Ahmed, Danielle Whicher, *Artificial Intelligence in Health Care: the Hope, the Hype, the Promise, the Peril*, vol. 2019, National Academies Press, 2022.
- [7] Hao Jiang, Yanning Zhou, Yi Lin, Ronald CK Chan, Jiang Liu, Hao Chen, Deep learning for computational cytology: A survey, *Med. Image Anal.* 84 (2023) 102691.
- [8] Kelei He, Chen Gan, Zhuoyuan Li, Islem Rekik, Zihao Yin, Wen Ji, Yang Gao, Qian Wang, Junfeng Zhang, Dinggang Shen, Transformers in medical image analysis, *Intell. Med.* 3 (1) (2023) 59–78.
- [9] Ling Zhang, Le Lu, Isabella Nogues, Ronald M Summers, Shaoxiong Liu, Jianhua Yao, DeepPap: deep convolutional networks for cervical cell classification, *IEEE J. Biomed. Health Inform.* 21 (6) (2017) 1633–1643.
- [10] Fahdi Kanavati, Naoki Hirose, Takahiro Ishii, Ayaka Fukuda, Shin Ichihara, Masayuki Tsuneki, A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images, *Cancers* 14 (5) (2022) 1159.
- [11] Wanli Liu, Chen Li, Ning Xu, Tao Jiang, Md Mamunur Rahaman, Hongzan Sun, Xiangchen Wu, Weiming Hu, Haoyuan Chen, Changhao Sun, et al., CVM-cervix: A hybrid cervical pap-smear image classification framework using CNN, visual transformer and multilayer perceptron, *Pattern Recognit.* 130 (2022) 108829.
- [12] Hans Ikenberg, Sabrina Lieder, André Ahr, Manfred Wilhelm, Christiane Schön, Arjola Khaja, Comparison of the hologic genius digital diagnostics system with the ThinPrep imaging system—a retrospective assessment, *Cancer Cytopathol.* 131 (7) (2023) 424–432.
- [13] Mariana T Rezende, Raniere Silva, Fagner de O Bernardo, Alessandra HG Tobias, Paulo HC Oliveira, Tales M Machado, Caio S Costa, Fatima NS Medeiros, Daniela M Ushizima, Claudia M Carneiro, et al., Cric searchable image database as a public platform for conventional pap smear cytology data, *Sci. Data* 8 (1) (2021) 151.
- [14] Jiawei Liu, Huijie Fan, Qiang Wang, Wentao Li, Yandong Tang, Danbo Wang, Mingyi Zhou, Li Chen, Local label point correction for edge detection of overlapping cervical cells, *Front. Neuroinform.* 16 (2022) 895290.
- [15] Balazs Harangi, Gergo Bogacsócs, Janos Toth, Ilona Kovacs, Erzsébet Dani, Andras Hajdu, Pixel-wise segmentation of cells in digitized pap smear images, *Sci. Data* 11 (1) (2024) 733.
- [16] David Kupas, Andras Hajdu, Ilona Kovacs, Zoltan Hargitai, Zita Szombathy, Balazs Harangi, Annotated pap cell images and smear slices for cell classification, *Sci. Data* 11 (1) (2024) 743.
- [17] Dávid Kupás, András Hajdu, Ilona Kovács, Zoltán Hargitai, Zita Szombathy, Balázs Harangi, APACC: Annotated pap cell images and smear slices for cell classification, 2023, <http://dx.doi.org/10.17605/OSF.IO/FP2XE>, <https://osf.io/fp2xe/>.
- [18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [19] Mingxing Tan, Quoc Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- [22] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V Le, Learning transferable architectures for scalable image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8697–8710.

- [23] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.
- [24] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, Lightgbm: A highly efficient gradient boosting decision tree, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [25] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama, Optuna: A next-generation hyperparameter optimization framework, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.
- [26] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, et al., Yolov10: Real-time end-to-end object detection, *Adv. Neural Inf. Process. Syst.* 37 (2024) 107984–108011.



David Kupas, Assistant Lecturer of AI & Data Science received his MSc degree in Computer Science in 2020. He is currently pursuing his Ph.D. and, since 2025, has been serving as an Assistant Lecturer at the Department of Data Science and Visualization. His research interests include medical image analysis, deep learning, and computer-aided diagnosis systems, with particular emphasis on automated cervical cancer screening and Pap-smear image analysis. He has authored or co-authored multiple peer-reviewed publications, and his work has received over 50 citations since 2020. His primary research fields are digital/medical image processing, pattern recognition, and machine learning/deep learning.



Ilona Kovacs is a consultant cytopathologist at the Department of Pathology, University of Debrecen Clinical Center, Kenézy Gyula Campus. She has decades of clinical experience in gynecological cytology and has played a leading role in the implementation and quality assurance of cervical cancer screening programs in the region.

Her current work includes the application of digital pathology and expert annotation of cytological samples for AI-assisted screening systems.



Andras Hajdu, Ph.D., Full Professor of AI & Data Science received an MSc degree in Mathematics from the University of Debrecen, Hungary, in 1996. He obtained his Ph.D. degree in Mathematics and Computer Science in 2003. He worked as a PostDoc researcher for the Artificial Intelligence Information Analysis Laboratory, Dept. of Informatics, Aristotle University of Thessaloniki between 2005–2006. Since 2017 he has been a full professor, since 2011 the head of the Department of Data Science and Visualization, and since 2019 the dean of the Faculty of Informatics, University of Debrecen. He is the leader of the local Data Science and Visualization doctoral program and the founder and director of the Gyorgy Hajos Data Science Students College. He worked as a data scientist for the Hungarian Data Asset Agency between 2019–2022. He serves as an instructor for the Microsoft Learn for Educators program and is a certified NVIDIA instructor and university ambassador. He has co-authored 57 journal papers and 120 conference papers cited more than 3500 times; his H-index is 29. His main interest lies in data science/artificial intelligence, medical image processing, and discrete mathematics.



Balazs Harangi, Ph.D., Associate Professor of AI & Data Science received his MSc degrees in Program Engineering and Mathematics from University of Debrecen, Faculty of Informatics, in 2010. From 2010 to 2013 he was a Ph.D. student at Faculty of Informatics and he obtained his Ph.D. degree in Informatics (Medical Image Processing) from the University of Debrecen, Hungary, in 2015. Since 2013 he has been serving as Assistant Lecturer, later as Assistant Professor and as Associate Professor at Department of Data Science and Visualization, Faculty of Informatics, University of Debrecen. He is a member of the IEEE, the John von Neumann Computer Society (Hungary) and member of the Hungarian Association for Image Analysis and Pattern Recognition. He has authored or co-authored 77 publications. His primary research fields are the digital/medical image processing, pattern recognition, machine learning/deep learning.