Journal of Intelligent Learning Systems and Applications, 2026, 18(3), 198-215
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

A Cervical Cell Classification Model: Assessing the Effect of Cellular Overlap

Suxiang Yu[1*], Jinyue Wu[2*], Dun Hua[1], Bai Yun[1], Huimiao Sun[1], Lingling Zhang[1], Feihong Wu[3], Dandan Yang[4], Xin Huang[5#]

[1]Department of Pathology, The Fourth Central Hospital of Baoding City, Baoding, China
[2]School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, China
[3]School of Design and Art, Shenyang Aerospace University, Shenyang, China
[4]Science and Education Department of Baoding Fourth Central Hospital, Baoding, China
[5]Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China
Email: 1197831790@qq.com, ywyjqyx@163.com, 15933556186@139.com, 1061047891@qq.com, 1600268443@qq.com, 3065187648@qq.com, wfh960125@163.com, bdsdszxyykjc@163.com, #huangxin@nbu.edu.cn

How to cite this paper: Yu, S.X., Wu, J.Y., Hua, D., Yun, B., Sun, H.M., Zhang, L.L., Wu, F.H., Yang, D.D. and Huang, X. (2026) A Cervical Cell Classification Model: Assessing the Effect of Cellular Overlap. *Journal of Intelligent Learning Systems and Applications*, **18**, 198-215.
https://doi.org/10.4236/jilsa.2026.183013

Received: March 10, 2026
Accepted: June 26, 2026
Published: June 29, 2026

Abstract

The prevalent overlapping phenomenon of cervical cells in clinical samples constitutes a primary barrier limiting the application of deep learning classification models in cervical cancer screening. This study aims to evaluate the adverse impact of cell overlapping on cell classification performance. Based on the YOLO11l-cls classifier, we fine-tuned the model using segmented non-overlapping single cells. The core approach involves comparing the model's performance on an ideal single-cell test set and a comparative over-lapping cell test set. Results show that the model exhibited outstanding performance under ideal conditions, achieving an AUC of 0.9959 and a precision of 0.9941. However, its performance declined significantly on the overlapping test set: the AUC dropped to 0.8816, and the precision decreased to 0.7368. These data strongly demonstrate the adverse effect of cell overlapping on cell classification, which leads to deterioration in the model's classification performance. Therefore, addressing the limitations in feature extraction of over-lapping cells is a critical prerequisite for improving the application value of classification models in complex clinical environments.

Keywords

Cervical Cancer, Cell Overlapping, YOLO11l-cls, Classification, Assessment

*These authors contributed equally to this work.
#Corresponding author.

1. Introduction

Cervical cancer is a major problem to the health of women in the world as it is the number one killer among women when it comes to cancer-related deaths in 37 countries and there is a disturbing trend towards an increase in young people contracting the disease [1] [2]. Timely screening and proper diagnosis of cervical cancer is very important to enhance prognosis since it can be cured entirely when diagnosed and treated early enough [3]. The prevention and control of cervical cancer mostly depends on pathological screening. Papanicolaou (Pap) smear and ThinPrep Cytologic Test (TCT) are critical tools in cervical cancer screening and they are aimed at determining and categorizing squamous intraepithelial lesions, allowing detection and treatment of pre-invasive stages of the disease before it advances to invasive phases [4]. Precision in the classification of cervical cytopathology images is an important step in the procedure. To ascertain the level of abnormality and evaluate whether the cancerous process is present, pathologists examine smear specimens using optical microscopy, which requires great skill and scrupulous examination to identify differences between normal and abnormal cells [5].

Classifying cells is a basic procedure in diagnosing diseases and biomedical studies especially during pathological evaluation on such diseases as cervical cancer, where correct characterization of malignancy grade of the cells is paramount. Cervical cell grading can be used to diagnose samples using morphological evaluation of cells present in samples to diagnose them in different grades including benign and atypical squamous cells through high-grade squamous intraepithelial lesions. Precise classifying is an essential basis that will be used to direct clinicians in coming up with proper treatment measures: lesions of low grade usually require no more than usual surveillance, but lesions of high grade should be subjected to urgent colposcopy or aggressive medical management. The accurate cell classification will allow high-risk patients to be treated in time and avoid unnecessary treatment of low-risk patients, which greatly contributes to the effective distribution of public health resources.

Conventional cell classification strategies have been based largely on manual identification of morphological properties combined with traditional machine learning methods like the support vector machines (SVM), decision trees, etc., to diagnose. Nonetheless, they require strong reliance on the quality of feature engineering and are prone to being unable to learn more intricate, nonlinear features specific to cellular morphology. The last few years have seen an explosion in the development of deep learning technologies and computer vision, which allows convolutional neural network-based models to make substantial progress in medical image analysis. These developments provide a powerful tool in the automated screening of cervical cancer that promises to be much more efficient and objective in clinical diagnosis [6].

Although the current breakthrough in deep learning has been impressive, cervical cells in clinical samples are usually presented as densely packed and with

substantial overlap [7]. At present, the majority of the work related to overlapping cell classification in the field of deep learning is divided into two major technical directions. The one that has gained a leading position is segmentation-based single-cell classification. Current studies and procedures to train models generally use non-overlapping images of individual cells taken out of high quality or sparse distributions of samples, allowing models to learn different morphological properties of individual cells efficiently. Although this allows the model to learn clear properties of the morphology, it does not take into account the fact that there are many overlapping cells in the real diagnostic setting, and how these can confuse classifiers. The pixel level features contained in overlapping areas are a composite of several cell instances, which makes it difficult to assign the features correctly to a certain cell with the help of classifiers. Therefore, important morphological properties of malignant cells can be hidden, which makes it more difficult to determine and identify such cells accurately and finally reduces the accuracy of their classification. The restriction considerably undermines the utility of deep learning models in clinical practice.

The purpose of this paper is to explore the influence of overlapping cells on cell classification performance in order to show that cell overlap presents a major problem to existing cell classification techniques. The research applied an advanced classification model, namely, YOLO11l-cls. Training inputs were single-cell images that had gone through segmentation and preprocessing to enable the model to learn how to correctly identify characteristics of typical single-cell lesions. In order to measure how much cell overlap impacts the effect, two different test sets were used to assess the performance of the model:

- Construction of comparison datasets: To obtain the actual cervical pathological images, two test sets were created, one with overlapping cells and the other one with clear, standardized non-overlapping cells that were extracted out of overlapping populations using accurate segmentation methods.
- Quantitative evaluation and conclusion: The pre-trained YOLO11l-cls classifier was optimized and then evaluated using the two image sets. The negative effect of cell overlap on classification accuracy was shown by comparing the important performance measures, such as area under the curve (AUC), accuracy and precision.

When applying the well-performing classifier in the practical clinical conditions, we have noticed a significant fact that when the model is used on a dataset with overlapping cells its performance drops significantly. Nevertheless, after eliminating overlapping cell cases in the dataset, the classification accuracy and precision recover immediately and substantially.

The current paper will be organized as follows: In Section 2, I review the literature concerning the automated classification of cervical cells and its solutions to cope with cell overlap. In Section 3, I focus on the origin of the dataset, the segmentation procedure, and how the two comparative test sets were created. The architecture of the core classifier, YOLO11l-cls, and the overall evaluation frame-

work are described in Section 4. The experimental setup, training hyperparameters, performance metrics and quantitative analysis of results are discussed in Section 5. Conclusively, Section 6 provides a summary of the negative effect of cell overlap on classification performance and concludes the study.

2. Related Works

The primary objective of automated analysis in cervical cancer cytopathology imaging is to achieve accurate classification of cells as either malignant or benign. Early computer-aided diagnosis (CAD) systems predominantly relied on machine learning approaches, including SVM [8], k-nearest neighbors (KNN) [9] [10], and random forests [11]. However, these methods exhibit a high dependence on handcrafted feature extraction and demonstrate limited generalization capability.

A significant proportion of high-performance cervical cell classification studies concentrate on achieving optimal diagnostic accuracy using ideal, feature-complete single-cell datasets, typically circumventing the challenges posed by overlapping morphological features through meticulous manual annotation. Xu *et al.* [12] developed a deep learning framework based on an enhanced Faster R-CNN architecture integrated with a shallow feature enhancement network and a generative adversarial network for cervical cell image classification. Their model achieved a maximum accuracy of 99.81% and an overall average accuracy of 89.4% across the SIPaKMeD and Herlev datasets. Rahaman *et al.* [13] introduced DeepCervix, an automated classification framework leveraging Hybrid Deep Feature Fusion (HDFF) technology, which demonstrated high classification accuracy in 2-class, 3-class, 5-class, and 7-class configurations on two publicly available datasets. Chen *et al.* [14] proposed an automated cervical cell screening system designed to support subsequent clinical diagnosis, with the CompactVGG classification model serving as its core component. The system achieved classification accuracies of 97.8% and 94.81% on the SIPaKMeD and Herlev datasets, respectively. Maurya *et al.* [15] presented a hybrid deep learning framework that employed transfer learning with Long Short-Term Memory (LSTM) and CNN models, as well as an ensemble approach combining Vision Transformer and CNN, trained and evaluated on the SIPaKMeD dataset, yielding high classification performance. Zhang *et al.* [16] proposed A2SDNet121, a deep convolutional neural network aimed at addressing the issue of low accuracy in cervical cell classification. The model was applied to binary and multi-class classification tasks on the Herlev and SIPaKMeD datasets, achieving an accuracy of up to 99%. Deepa *et al.* [17] adopted a two-stage approach involving segmentation followed by classification, utilizing a CNN model to classify overlapping cervical smear cells. Overlapping cells were segmented using the midpoint algorithm, and the cropped images of individual cells were then fed into the model for distinguishing between normal and abnormal cells, resulting in a classification accuracy of 96%.

To address the challenges posed by cell clustering and overlapping, researchers have primarily employed object detection or instance segmentation techniques.

Furthermore, only a limited number of studies have attempted to classify overlapping cells directly without relying on prior segmentation. Mulmule *et al.* [18] introduced a fully automated transfer learning framework that extracts 128 × 128 pixel patches containing multiple cervical cells from high-resolution images, enabling direct classification of overlapping cells under microscopic examination with a reported accuracy of 99.86%. Nguyen *et al.* [19] proposed a method for segmenting and classifying overlapping cervical cells based on semi-supervised learning, utilizing the SemiFixMatch model and achieving competitive performance in this task. In another study, Mulmule *et al.* [20] fine-tuned and retrained AlexNet along with pre-trained models from ImageNet and Places365 on the CERVIX93 dataset using cervical cancer enhancement data, evaluating model performance across several metrics. Their results demonstrated that the AlexNet model achieved the highest predictive performance in cervical cancer prediction, with an accuracy rate of 99.03%.

The main related work of cervical cell image classification are summarized in Table 1.

Table 1. Work related to cervical cell classification.

No.	Data	Features	Methods/Model	Cell Input Type	Performance	Ref.
1	Pap-smear images dataset	morphologica and textual features of the cells	SVM	Single	Accuracy (2-class): 98.83%	[8]
2	Pap-smear images Dataset collected from Fortis Hospital Mohali, Punjab (India)	morphological features of the cells	k-nearest neighbors	Single	Accuracy: 82.9%	[9]
3	Pap smear images from the Herlev Database of Danish Hospitals	geometric feature extraction	k-nearest neighbors	Single	Accuracy: 94.29%	[10]
4	single Pap cell images	11 nucleus features and 9 cytoplasm features	Random forest	Single	AUC (13 features): 98.04%	[11]
5	SIPaKMeD and Herlev Pap smear datasets	Deep learning features	Faster R-CNN/Shallow Feature Enhancement/GAN	Single	Max Acc: 99.81%	[12]
6	SIPaKMeD Pap smear dataset	Deep learning features	CNN and ensemble learning	Single	Accuracy (2-class): 99.81%	[13]
7	SIPaKMeD and Herlev Pap-smear images dataset	Deep learning features	CompactVGG model	Single	Acc (SIPaKMeD): 97.80%, Acc (Herlev): 94.81%	[14]
8	SIPaKMeD Pap smear dataset	Deep learning features	Vit-CNN, CNN-LSTM	Single	Acc (Vit-CNN): 97.65% Acc (CNN-LSTM): 95.80%	[15]
9	SIPaKMeD and Herlev Pap smear datasets	Deep learning features	A2SDNet121 model	Single	Acc (SIPaKMeD): 99.22% Acc (Herlev): 99.14%	[16]

Continued

10	Herlev Pap smear images dataset	Deep learning features	CNN with reLU and Mid-point segmentation Algorithm	Single	Accuracy: 96%	[17]
11	Cervix 93 cervical cytology dataset	Deep learning features	Alexnet	Overlap	Accuracy: 99.86%	[18]
12	ISBI2014 Pap smear dataset	Deep learning features	SemiFixMatch model	Single	model total loss: 0.08926	[19]
13	Cervix 93 cervical cytology dataset	Deep learning features	Alexnet, ImageNet, Places365 model	Overlap	Alexnet Acc: 99.03%	[20]
14	Pap smear dataset collected from the fourth central hospital of Baoding city, China	Deep learning features	YOLO11l-cls	Single/ overlap	Acc: Single 94.06%/Overlap 83.05%	This Work

3. Data

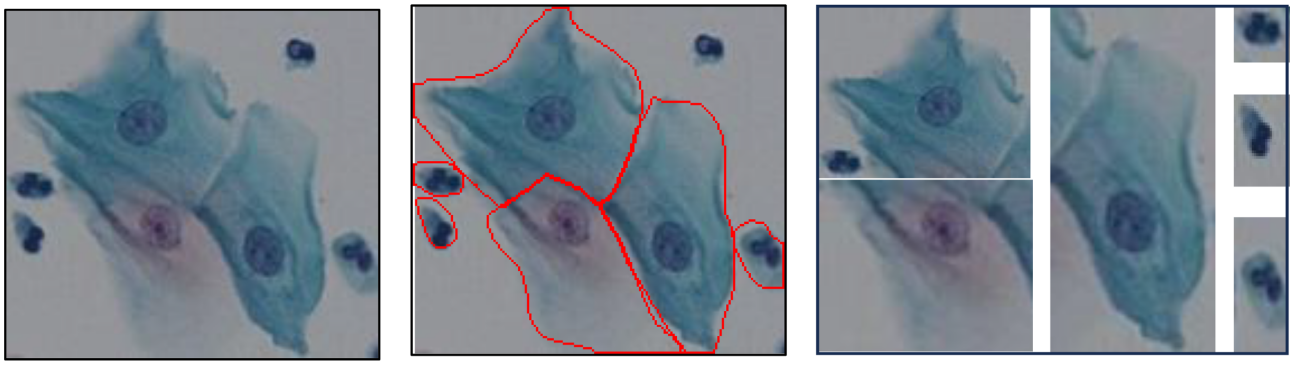
This study was conducted using 2504 overlapping cervical cell images obtained from the Fourth Central Hospital of Baoding, and received formal approval from the hospital's Institutional Review Board prior to commencement.

The pathological images of cervical cancer cells used in the current research have been obtained based on the data set reported in the paper by Yu and Feng *et al.* [21] which originated at the Fourth Central Hospital of Baoding, China. This dataset consists of 2504 overlapping cervical cell images. In order to make the classification model learn correct morphologic characteristics of cellular lesions the study uses high-quality and non-overlapping cell instances as inputs to the classifier when training. To achieve this, we use the developed Cellpose-SAM [22] deep learning based instance segmentation model as the main method in constructing the datasets. Cellpose-SAM is a segmentation model that combines the accurate instance separation approach of Cellpose [23] with the strong general feature learning properties of SAM, which has shown better generalization results and segmentation quality when morphologically complex and densely packed cells are processed. It allows extracting individual cell instances of the original images reliably, thus forming a sound basis of model training.

Figure 1 and Figure 2 show the areas of overlap between normal and abnormal cervical tissues in this dataset as well as non-overlapping cell maps produced by instance segmentation, respectively. The normal cervical cells have a central nucleus and a cytoplasm which can be easily differentiated against the background, where the nuclear region is less than the cytoplasmic region. Conversely, the cancerous cervical cells tend to have fibrous cytoplasm and enlarged nuclei.

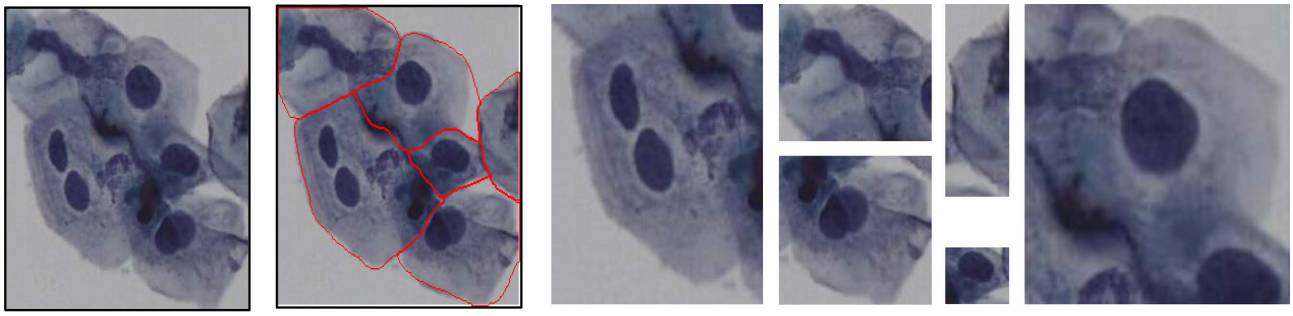
From an initial dataset of 2504 images, 433 images exhibiting cell overlap were selected. These images were segmented using the CellPose-SAM model, generating 4624 non-overlapping cell samples. The segmented samples were classified based on their cellular content: a sample was labeled as abnormal if it contained any abnormal

cells, and normal otherwise. Consequently, 3130 normal and 1494 abnormal samples were obtained. The detailed dataset composition is summarized in Table 2.



(a) (b) (c)

Figure 1. Images of normal cervical cells. (a) represents the original image; (b) displays the ground truth derived from manual segmentation; (c) presents the instance map containing six individually cropped cells based on segmentation results.



(a) (b) (c)

Figure 2. Images of abnormal cervical cells. (a) represents the original image; (b) displays the ground truth derived from manual segmentation; (c) presents the instance map containing six individually cropped cells based on segmentation results.

Table 2. Summary of data characteristics in this study.

	Normal Cell	Abnormal Cell	Total
Overlapping Cell	376	57	433
Single Cell	3130	1494	4624
Total	3606	1551	5057

4. Method

4.1. Overview of the Proposed Approach

In selecting the classification model, this study prioritized both architectural advancement and feature extraction efficacy, ultimately employing YOLO11l-cls—a pre-trained model from the YOLOv11 [24] series by Ultralytics—as the core classifier. As a state-of-the-art visual architecture, YOLO11 demonstrates significant advantages in global feature aggregation and computational efficiency. The specific selection of the "large" version was tailored to the complexity of the dataset. Compared to lightweight variants, YOLO11l features a deeper network and

a larger parameter capacity, enabling it to more acutely capture fine-grained morphological differences. Conversely, compared to the extra-large version, it effectively controls computational costs and mitigates overfitting risks while maintaining high accuracy, thereby achieving an optimal balance between performance and efficiency.

The overall methodology is illustrated in Figure 3. The non-overlapping cervical cell dataset, obtained through segmentation, was partitioned into training and testing sets. The training set was utilized to fine-tune the pre-trained YOLO11l-cls model, enabling it to adapt to the pathological characteristics of cervical cells. To quantitatively evaluate the impact of cell overlap on classification performance, comparative test sets featuring overlapping cells were introduced. Subsequently, the fine-tuned classifier was applied to these test sets, and key performance metrics—including accuracy—were computed to empirically analyze the influence of cellular overlap.

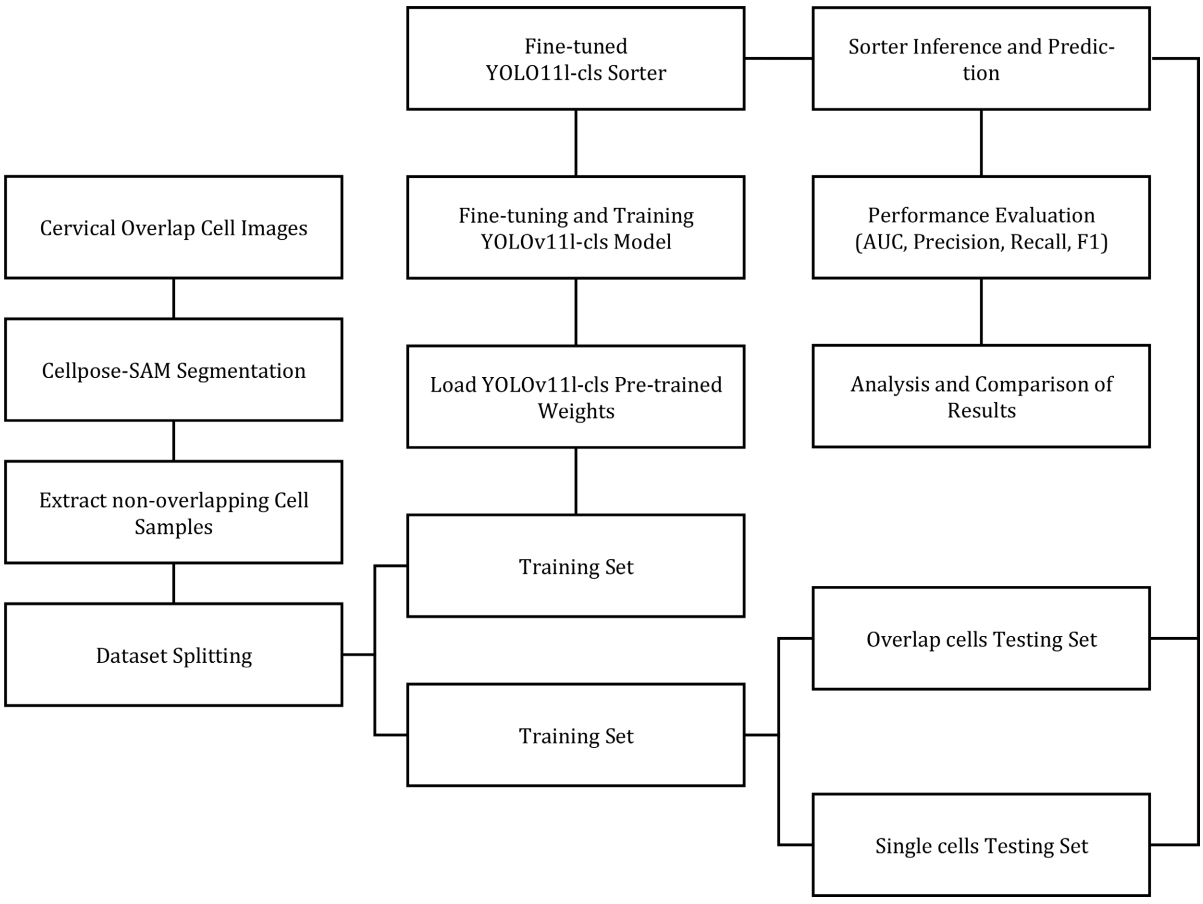


Figure 3. Flowchart of the overall methodology.

4.2. Overview of the YOLO11l-cls Architecture

YOLO [25] is one of the efficient object detection algorithms created by Joseph Redmon *et al.* It is based on a single stage framework which concurrently estimates bounding boxes and classes using a single forward through the whole image

[26]. YOLOv11 is a new development in the YOLO series aimed at improving the effectiveness and accuracy in feature extraction. Its classification version, YOLO11l-cls, has also been improved and is specifically trained to perform image classification. The architecture is based on the well-known three-stage structure of the YOLO series: backbone network, neck and head, and it provides an end-to-end workflow, including feature initialization, deep feature extraction, attention enhancement, and classification output. Figure 4 displays the main architectural elements of YOLO11l-cls.

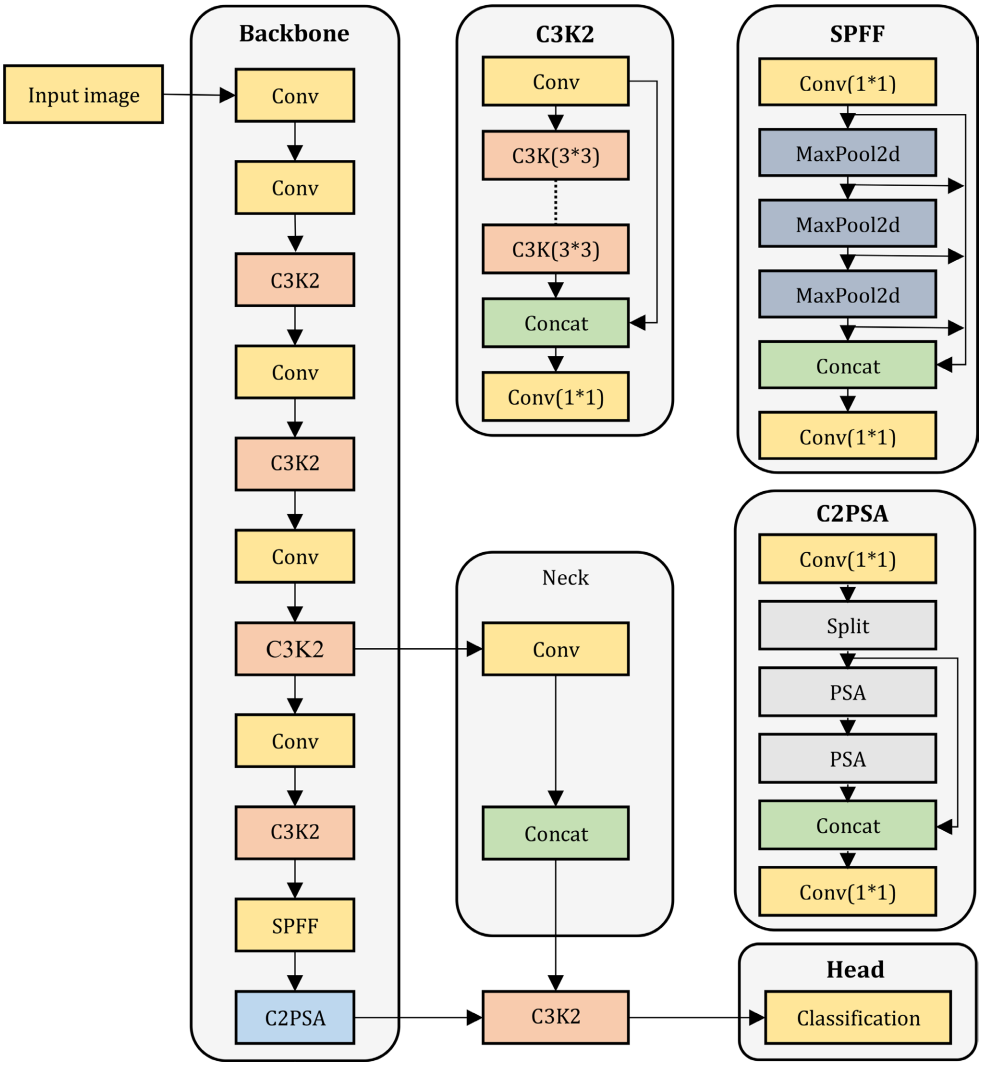


Figure 4. Key architectural modules of YOLO11l-cls, adapted from Rasheed's paper [27].

Backbone network is the main part of the model that performs initial feature extraction and spatial down sampling. It mainly contains traditional convolutional layers with a few highly effective modules. The C3K2 module acts as the hub of deep features extraction and inter-stage local connection with its aim to reduce redundant computations without losing high level feature representations [25]. In particular, the input feature map is divided according to the channel di-

mension: one branch is routed into a subnetwork of N Bottleneck architectures, where each Bottleneck uses two downsized convolutional kernels to get a lightweight architecture; the other branch serves as a skip connection. Both branches are finally concatenated and fused, thus resulting in an optimal trade-off between computational efficiency and detection accuracy. Moreover, the backbone keeps the SPPF module, which is a computationally efficient version of spatial pyramid pooling. With three consecutive applications of Max Pooling layers with the same kernel size on the aggregation of the multi-scale spatial context, the outputs are combined across stage, improving the model resilience to changes in the input image scale. The C2PSA module is introduced as one of the main architectural innovations in YOLOv11, which allows attention-based improvement. The module incorporates a lightweight self-attention mechanism into the Bottleneck stack on the C2P trunk path that allows the model to dynamically recalibrate feature responses by assigning adaptive weights to various channels or spatial positions. This enhances greatly the ability of the model to concentrate on discriminative microscopic features, e.g., the cervical nucleus and cytoplasm.

Neck module is a multi-scale feature integrator that transmits multi-scale features to the head to predict, using an effective C3K2 block to improve the total performance of feature aggregation. The C2PSA attention system goes even further to narrow down the spatial attention by focusing on informative region. This design can be used in cases where small or poorly defined features are involved, like cervical cell images, where this design makes it easy to integrate and add value to context information with high levels of discrimination such that the classification head sees highly discriminative feature representations.

The detection head produces the last prediction results, with C3K2 blocks in the pathway processing of multi-scale features at different depths. The next step is the refinement of the feature maps by a CBS layer, and these maps are passed through both global average pooling layer and a fully connected layer to obtain the final class scores. This model architecture allows the model to successfully map high-dimensional image features and output them as probabilistic outcomes on binary classification, thus being able to accurately classify the end-to-end images.

4.3. Classification Prediction Process

The YOLOv11 offers various different pre-trained models, and the YOLO111-cls demonstrates a very high Top-1 rate on one of the standard image recognition benchmarks, ImageNet, achieving 78.3 percent at 224×224 input size, which confirms its strong ability to extract features and the effective construction of its architecture. Consequently, YOLO11l-cls was chosen as the main classifier of this research and certain parameters of the YOLO11l-cls model were optimized and retrained in order to make its feature representation properly fit the particular pathological features of cervical cell images, thus optimizing its classification performance in medical imaging tasks.

Within this work, the YOLO11l-cls model views cervical cell binary classifica-

tion as a typical image classification task with a flow-through end-to-end design that will predict class probabilities of the input single-cell images directly. The classification pipeline, depicted in Figure 5, contains several important steps, including:

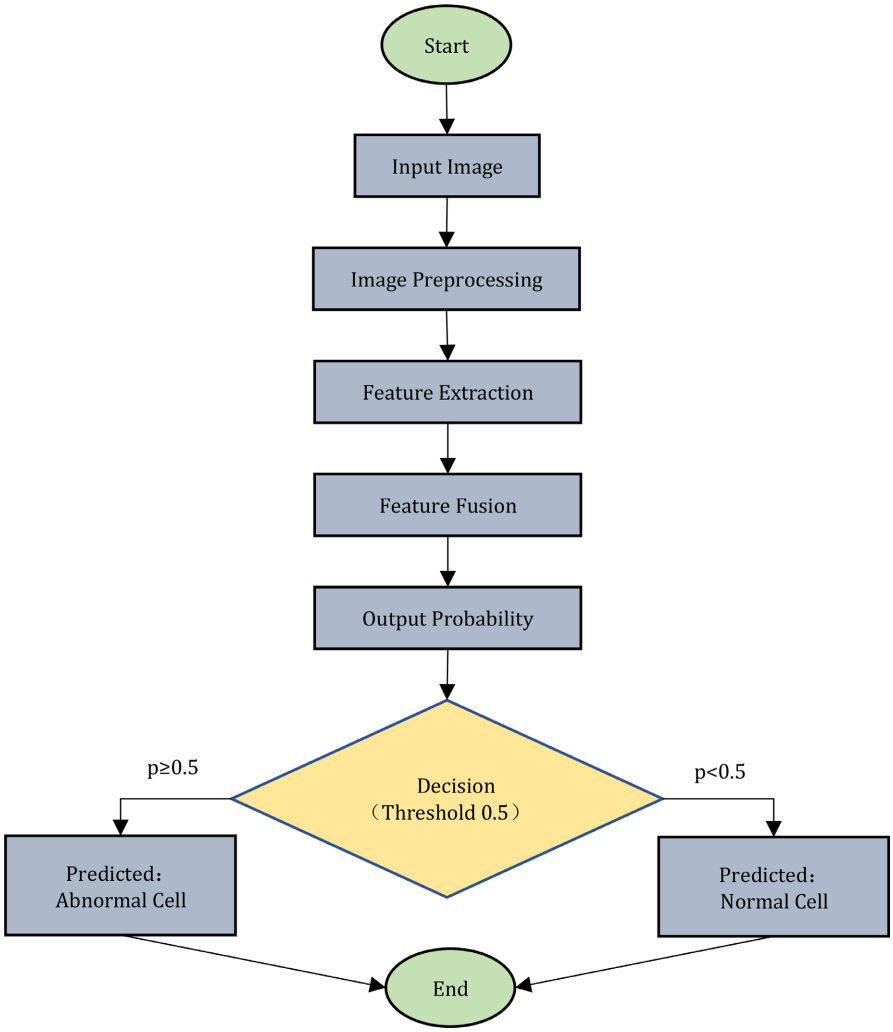


Figure 5. Flowchart of the classification prediction process.

1) Image preprocessing and input: Every cell image is rescaled to the dimensions that the model expects as an input, and it is also subjected to data augmentation methods. Then the preprocessed image is used as an input to the backbone network of the model.

2) Feature Extraction and Dimensionality Reduction: The backbone network converts raw pixel information into high-dimensional semantic feature maps through several convolutional and pooling layers. With increasing depth, the spatial dimensions of the feature maps decrease progressively.

3) Global Feature Aggregation: The final feature map is subjected to a global average pooling operation which produces a compact and fixed-length feature vector that retains the most salient features.

4) Final Classification Prediction: The aggregated feature vector is passed through a fully connected layer, which projects it into the output class space. For binary classification, a sigmoid activation function is applied to produce a probability score $P \in [0, 1]$. If $P \geq 0.5$, the cell is classified as abnormal; otherwise, it is classified as normal.

Regarding the classification probability threshold, this study adopted the default value of 0.5 as the decision boundary. Because 0.5 serves as the natural decision boundary for the terminal Sigmoid activation functions, preserving it maximizes the reflection of the original posterior probability distribution learned by the network and prevents potential overfitting to a specific validation set.

5. Experiment and Results

5.1. Dataset Splitting

To support the quantitative analysis of how overlapping cells affect cervical cytology images, the 4624 non-overlapping cell instances were partitioned into a training set (3321 instances) and a test set (1303 instances) at an approximate 7:3 ratio. To rigorously prevent data leakage, this split was performed strictly at the whole-slide image patch level, ensuring that cell instances originating from the same original field of view were mutually exclusive between the training and test datasets. Furthermore, three distinct comparative test sets were constructed to evaluate theoretical performance and practical applicability across varying degrees of complexity:

Test Set 1: Comprises the 1303 non-overlapping cell instances, designed to assess the model's theoretical baseline performance under ideal imaging conditions.

Test Set 2: Consists of 188 original overlapping cell images supplemented by a 20% subset of non-overlapping cell instances (n = 260) drawn from the benchmark set. This hybrid setup aims to evaluate the model's classification robustness in complex, clinically realistic scenarios.

Test Set 3: Contains exclusively the 188 original overlapping cell images, established to specifically isolate and assess the model's diagnostic efficacy when confronting purely overlapping cellular structures.

5.2. Data Preprocessing and Augmentation

In order to make the model more robust and able to extrapolate to other geometrical modifications of cells, and changes in lighting conditions in the cell classification problem in the cervix, all input images are rescaled to 640×640 pixels in size before training. The training stage is conducted using a strict and focused data augmentation approach that includes horizontal flipping, scaling, and random change in brightness. In particular, a random translation with an offset parameter of 0.5 is used to mimic the natural spatial configuration of cells within the field of view and reduce the effect of boundary artifacts. To enhance data variety, horizontal flipping is done at probability of 0.5, and the original image orientation is maintained in case no flip has been done. Random brightness transformations are used to simulate changes in the lighting and staining conditions, allowing the

model to adjust to a wide range of imaging conditions. All data augmentation processes were done through the Ultralytics library on data loading.

5.3. Model Loading and Training

To classify the segmented overlapping cervical cytopathology images, the YOLO11l-cls model was initialized with pre-trained weights and subsequently fine-tuned. To maximize classification performance, several key training configurations were adjusted. First, the input image resolution was increased to preserve fine-grained cellular morphological features. For model optimization, the Adam optimizer was employed to accelerate convergence on the limited medical dataset, paired with an appropriately tuned learning rate to ensure stable weight updates and optimal loss minimization. To mitigate potential overfitting, weight decay was moderately increased, and a 20% dropout rate was introduced. Furthermore, the training was conducted over 500 epochs with a random seed of 42 to ensure reproducibility, and an early stopping patience of 100 epochs was applied. A comprehensive summary of all experimental parameters is provided in Table 3.

Table 3. Experimental parameter settings.

Parameter	Default Value	Setting
imgsz	224	640
lr0	0.2	0.001
cos_lr	True	False
momentum	0.9	0.937
optimizer	SGD	Adam
batch Size	256	16
weight_decay	0.0001	0.0005
drop out	0.0	0.2
hsv_s	0.4	0.7

5.4. Performance Metrics

After the YOLO11l-cls model was trained, this paper produced the receiver operating characteristic (ROC) curve to assess the performance of the model in terms of classification. The ROC curve is a popular method of evaluating binary classification models and shows the trade-off between the true positive rate and the false positive rate at different classification thresholds. Being a one-dimensional measurement, the AUC number—which is closest to 1.0—is often considered a sign of better overall performance of the model and can be used to compare the generalization abilities of the model with various threshold values.

Moreover, the accuracy, precision, recall, and the F1 score were also used to assess the performance of classification. These measures are based on four basic results, *i.e.*, true positives (TP), false positives (FP), true negatives (TN), and false

negatives (FN). Precisely, TP indicates the situations where the abnormal cells are properly categorized as abnormal; FP means the situation when normal cells are wrongly classified as abnormal; TN means the case where normal cells are properly classified as normal; and FN means the abnormal cells that are wrongly predicted as normal.

Accuracy measures how accurately the model detects abnormal cells, *i.e.*, the proportion of actually abnormal specimens out of all those that are predicted to be abnormal. It can be calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Recall reflects the capacity of the model to recognize abnormal cells and is expressed as the percentage of true abnormal samples that are correctly predicted. The formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall measures the model's ability to identify abnormal cells and represents the proportion of actual abnormal samples that are correctly predicted. It is calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1 score is a harmonic mean of precision and recall to give an overall picture of the performance of a model. The equation is:

$$\text{F1 score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

5.5. Analysis of Results

Figures 6-8 display the ROC curves for Test Sets 1, 2, and 3, respectively, with the

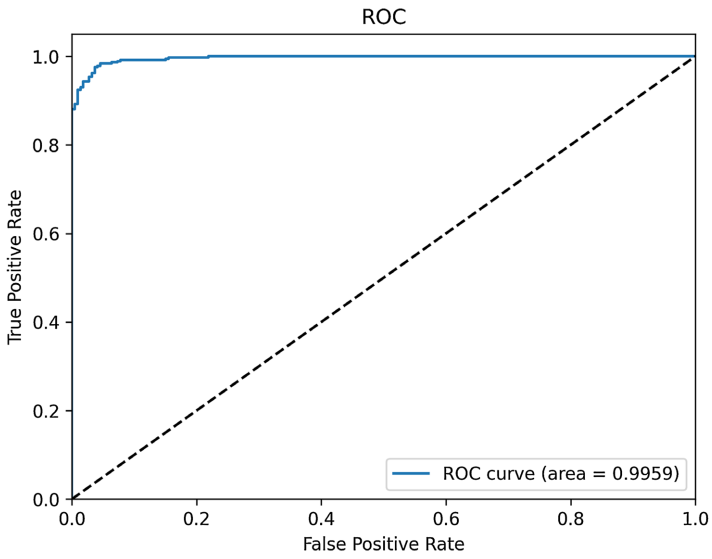


Figure 6. ROC curve for test set 1.

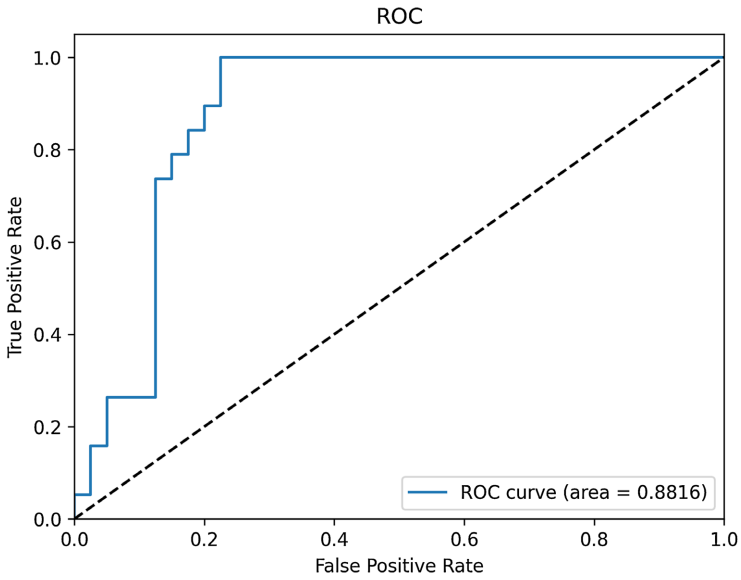


Figure 7. ROC curve for test set 2.

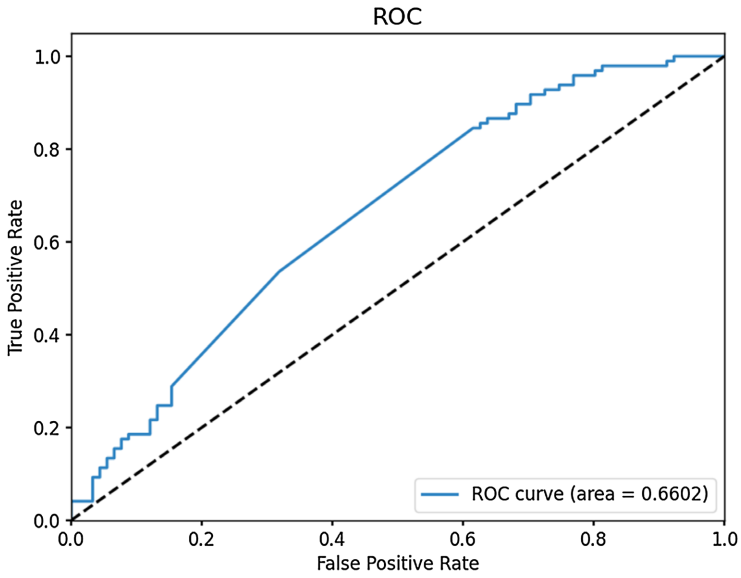


Figure 8. ROC curve for test set 3.

Table 4. Performance metrics.

	AUC	Accuracy	Precision	Recall	F1 score
Test Set 1	0.9959	0.9406	0.9941	0.9108	0.9255
Test Set 2	0.8816	0.8305	0.7368	0.7368	0.7809
Test Set3	0.6602	0.6064	0.6420	0.5361	0.5843

associated performance metrics summarized in Table 4. The ROC curve for Test Set 1 is positioned closest to the upper-left corner of the plot, reflecting superior classification performance with AUC of 0.9959. Conversely, a noticeable decline in performance is observed when the model is evaluated on Test Set 2. The AUC

decreases to 0.8816, representing an approximate 11.5% drop from the baseline. Furthermore, the accuracy, precision, recall, and F1 score for Test Set 2 fall to 0.8305, 0.7368, 0.7368, and 0.7809, respectively. This performance degradation is most pronounced in Test Set 3, where the AUC further drops to 0.6602, accompanied by an accuracy of 0.6064, precision of 0.6420, recall of 0.5361, and an F1 score of 0.5843.

6. Conclusions

In this study, the YOLO11l-cls model was employed to systematically investigate and quantify the adverse effects of cellular overlap on cervical cytology classification through comparative experiments. Trained exclusively on segmented, morphologically distinct single-cell data and evaluated on an ideal baseline of non-overlapping cell instances (Test Set 1), the model demonstrated near-optimal theoretical diagnostic accuracy. This was evidenced by an AUC of 0.9959, an F1-score of 0.9255, and a remarkably high precision of 0.9941, underscoring its robust feature extraction capabilities and high diagnostic consistency.

However, when the trained classifier was applied to a mixture of overlapping and non-overlapping cells (Test Set 2), significant degradations across all evaluation metrics were observed. Compared to the baseline, the AUC decreased to 0.8816, representing an absolute reduction of 0.1143 (an approximate 11.5% drop). Concurrently, precision fell by 0.2573 (from 0.9941 to 0.7368), recall decreased by 0.1740 (from 0.9108 to 0.7368), and the F1-score dropped by 0.1446 (from 0.9255 to 0.7809). Crucially, the overall classification accuracy was reduced to 0.8305.

This performance deterioration became most acute in Test Set 3, which consists exclusively of overlapping cell images. Under these extreme conditions, the model's diagnostic capacity suffered a severe bottleneck, with the AUC plummeting to 0.6602, accompanied by a sharp decline in accuracy (0.6064), precision (0.6420), recall (0.5361), and an F1-score of 0.5843.

This stepwise performance decay robustly confirms that cellular overlap poses a critical barrier to automated models achieving their theoretical performance limits. The pronounced drops in precision and recall reveal that overlapping cell structures severely impair the network's ability to recognize and interpret pivotal diagnostic features, particularly nuclear morphology. Automated cervical cytology classification systems remain heavily reliant on high-quality, morphologically intact cellular inputs. Therefore, it is imperative to address the challenges of instance segmentation and feature extraction in the presence of overlapping cells to enhance the real-world applicability and clinical utility of these models.

Funding

This research was supported by the S&T Program of Hebei under project No. 22377774D. This grant ensured the smooth implementation of this research, and we would like to express our heartfelt gratitude.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., *et al.* (2024) Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **74**, 229-263. https://doi.org/10.3322/caac.21834

[2] Burmeister, C.A., Khan, S.F., Schäfer, G., Mbatani, N., Adams, T., Moodley, J., *et al.* (2022) Cervical Cancer Therapies: Current Challenges and Future Perspectives. *Tumour Virus Research*, **13**, Article ID: 200238. https://doi.org/10.1016/j.tvr.2022.200238

[3] Bouvard, V., Wentzensen, N., Mackie, A., Berkhof, J., Brotherton, J., Giorgi-Rossi, P., *et al.* (2022) The IARC Perspective on Cervical Cancer Screening. *Obstetrical & Gynecological Survey*, **77**, 154-156. https://doi.org/10.1097/ogx.0000000000001017

[4] Samikhovna, M.K. (2025) Modern Understanding of the Diagnosis and Prevention of Cervical Cancer. *Modern Education and Development*, **19**, 339-349.

[5] Staats, P.N., Souers, R.J., Nunez, A.L., Li, Z., Kurtycz, D.F.I., Goodrich, K., *et al.* (2020) The Differential Diagnosis of Reparative Changes and Malignancy: Performance in the College of American Pathologists Pap Education and Proficiency Testing Programs. *Archives of Pathology & Laboratory Medicine*, **144**, 846-852. https://doi.org/10.5858/arpa.2019-0298-cp

[6] Youneszade, N., Marjani, M. and Pei, C.P. (2023) Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges. *IEEE Access*, **11**, 6133-6149. https://doi.org/10.1109/access.2023.3235833

[7] Chen, E., Ting, H., Huang Chuah, J. and Zhao, J. (2024) Segmentation of Overlapping Cells in Cervical Cytology Images: A Survey. *IEEE Access*, **12**, 114170-114189. https://doi.org/10.1109/access.2024.3445371

[8] Chen, Y., Huang, P., Lin, K., Lin, H., Wang, L., Cheng, C., *et al.* (2014) Semi-automatic Segmentation and Classification of Pap Smear Cells. *IEEE Journal of Biomedical and Health Informatics*, **18**, 94-108. https://doi.org/10.1109/jbhi.2013.2250984

[9] Sharma, M., Kumar Singh, S., Agrawal, P. and Madaan, V. (2016) Classification of Clinical Dataset of Cervical Cancer Using KNN. *Indian Journal of Science and Technology*, **9**, 1-5. https://doi.org/10.17485/ijst/2016/v9i28/98380

[10] Kusumawardani., L.A., Rulaningtyas, R. and Winarno, W. (2023) Classification of Cervical Cancer Cells Using the K-Nearest Neighbor (KNN) Method Based on Geometric Feature Extraction. *AIP Conference Proceedings*, **2858**, Article ID: 030003. https://doi.org/10.1063/5.0167165

[11] Sun, G. (2017) Cervical Cancer Diagnosis Based on Random Forest. *International Journal of Performability Engineering*, **13**, 446-457. https://doi.org/10.23940/ijpe.17.04.p12.446457

[12] Xu, L., Cai, F., Fu, Y. and Liu, Q. (2023) Cervical Cell Classification with Deep-Learning Algorithms. *Medical & Biological Engineering & Computing*, **61**, 821-833. https://doi.org/10.1007/s11517-022-02745-3

[13] Rahaman, M.M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., *et al.* (2021) DeepCervix: A Deep Learning-Based Framework for the Classification of Cervical Cells Using Hybrid Deep Feature Fusion Techniques. *Computers in Biology and Medicine*, **136**, Ar-

ticle ID: 104649. https://doi.org/10.1016/j.compbiomed.2021.104649

[14] Chen, H., Liu, J., Wen, Q., Zuo, Z., Liu, J., Feng, J., *et al.* (2021) CytoBrain: Cervical Cancer Screening System Based on Deep Learning Technology. *Journal of Computer Science and Technology*, **36**, 347-360. https://doi.org/10.1007/s11390-021-0849-3

[15] Maurya, R., Nath Pandey, N. and Kishore Dutta, M. (2023) VisionCervix: Papanicolaou Cervical Smears Classification Using Novel CNN-Vision Ensemble Approach. *Biomedical Signal Processing and Control*, **79**, Article ID: 104156. https://doi.org/10.1016/j.bspc.2022.104156

[16] Zhang, Y., Ning, C. and Yang, W. (2025) An Automatic Cervical Cell Classification Model Based on Improved DenseNet121. *Scientific Reports*, **15**, Article No. 3240. https://doi.org/10.1038/s41598-025-87953-1

[17] Deepa, T.P. and Rao, A.N. (2024) Classification of Normal and Abnormal Overlapped Squamous Cells in Pap Smear Image. *International Journal of System Assurance Engineering and Management*, **15**, 519-531. https://doi.org/10.1007/s13198-022-01805-z

[18] Mulmule, P.V. and Kanphade, R.D. (2021) Classification of Overlapping Cells in Microscopic Cervical Images: A Transfer Learning Approach. 2021 *Asian Conference on Innovation in Technology* (*ASIANCON*), Pune, 27-29 August 2021, 1-7. https://doi.org/10.1109/asiancon51346.2021.9544587

[19] Nguyen, D.H. and Nguyen, T.N.L. (2025) A Study on Semi-Supervised Solutions for Overlapped Cell Classification. Vietnam-Korea University of Information and Communication Technology, Working Paper. https://elib.vku.udn.vn/handle/123456789/5009

[20] Mulmule, P.V. and Kanphade, R.D. (2022) Classification of Cervical Cytology Overlapping Cell Images with Transfer Learning Architectures. *Biomedical and Pharmacology Journal*, **15**, 277-284. https://doi.org/10.13005/bpj/2364

[21] Yu, S., Feng, X., Wang, B., Dun, H., Zhang, S., Zhang, R., *et al.* (2021) Automatic Classification of Cervical Cells Using Deep Learning Method. *IEEE Access*, **9**, 32559-32568. https://doi.org/10.1109/access.2021.3060447

[22] Pachitariu, M., Rariden, M. and Stringer, C. (2025) Cellpose-SAM: Superhuman Generalization for Cellular Segmentation. bioRxiv. https://doi.org/10.1101/2025.04.28.651001

[23] Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M. (2021) Cellpose: A Generalist Algorithm for Cellular Segmentation. *Nature Methods*, **18**, 100-106. https://doi.org/10.1038/s41592-020-01018-x

[24] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. arXiv: 2410.17725.

[25] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[26] Jegham, N., Koh, C.Y., Abdelatti, M. and Hendawi, A. (2024) YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. arXiv: 2411.00201v3.

[27] Rasheed, A.F. and Zarkoosh, M. (2025) YOLOv11 Optimization for Efficient Resource Utilization. *The Journal of Supercomputing*, **81**, Article No. 1085. https://doi.org/10.1007/s11227-025-07520-3