

CervixNet: An Overlap-Aware Multi-Task Deep Learning Framework for Whole-Slide Cervical Cytology Classification

Shubhaker B ^{1,✉}, Raghavendra B.S. ¹

¹. Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, IND

Received: April 20, 2026 | Review began: May 04, 2026 | Review ended: May 18, 2026 | Published: June 02, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Automated analysis of cervical Papanicolaou smear slides faces a practical obstacle that most published systems bypass: cells in real clinical preparations do not sit in isolation. They overlap, stack, and cluster. Every prior deep learning approach evaluated on the SIPaKMeD benchmark was trained on clean single-cell crops, rendering those systems unsuitable for deployment on whole-slide images (WSIs) where cell clutter is the norm. This paper presents CervixNet, an end-to-end framework that tackles overlapping-cell classification through three coordinated contributions: (1) Overlap Copy-Paste Augmentation (OCA), an on-the-fly synthesis strategy using Poisson blending that generates realistic multi-cell training fields without additional annotation; (2) Model-C, a multi-task architecture pairing an EfficientNet-B0 backbone with a squeeze-and-excitation channel attention module and an auxiliary U-Net segmentation decoder; and (3) a WSI Overlap Detector (Model-B), a lightweight binary convolutional network that screens tile-level exports from WSIs and routes overlap-rich regions to the classifier. On the SIPaKMeD five-class benchmark ($N = 1,505$ test images), CervixNet achieves 90.70% accuracy, a macro F1 of 0.9079, and a macro receiver operating characteristic area under the curve of 0.987. Error analysis stratified by overlap score reveals no misclassification for cells with overlap scores above 0.4 in the test set, validating OCA's capacity to substantially reduce high-overlap failures. Training and validation curves confirm stable convergence with minimal overfitting.

Categories: Image Processing and Analysis, Computer Vision, Deep Learning

Keywords: cervical cytology, pap smear, whole-slide image, overlapping cell, efficientnet, multi-task learning, u-net, sipakmed, deep learning

Introduction

Cervical cancer is the fourth most prevalent malignancy in women globally, accounting for roughly 340,000 deaths each year [1]. The Papanicolaou (Pap) smear remains the standard screening method in most health systems, yet its diagnostic interpretation depends on trained cytologists who must examine thousands of cells per slide - a process that is slow, costly, subject to inter-observer variability, and often inaccessible in resource-limited settings [1].

Deep learning approaches have accelerated progress on automated Pap smear classification since 2016. The SIPaKMeD dataset [2], introduced by Plissiti et al. in 2018 with five labeled classes of cervical squamous cells, has become the standard benchmark. Results on it, however, conceal a clinically important gap: every published system was trained on individual cell crops extracted under controlled conditions [3]. When these classifiers encounter whole-slide images (WSIs) - where cells stack, overlap, and cluster - performance degrades substantially and without warning [4].

How to cite this article:

B S, B.S. R (June 02, 2026) CervixNet: An Overlap-Aware Multi-Task Deep Learning Framework for Whole-Slide Cervical Cytology Classification. Cureus J Comput Sci 3 : es44389-026-00095-x. DOI https://doi.org/10.7759/s44389-026-00095-x

The overlap problem is not marginal. Koilocytes - human papilloma virus (HPV)-marker cells and the most diagnostically important early-stage finding - show variable perinuclear halo morphology that becomes ambiguous when neighboring cells intrude on the field [2]. A systematic review by Jiang et al. [3] identified cell clutter as a primary unresolved challenge, noting that existing pipelines universally assume clean single-cell inputs at the classification stage. No published automatic system provides a principled mechanism for handling this. CervixNet fills that gap through a training strategy designed specifically to improve robustness to overlap.

Prior deep learning work on the SIPaKMeD benchmark spans classical convolutional approaches [5], ensemble convolutional neural networks (CNNs) [6], and more recent transformer-based frameworks [7,8]. A comprehensive survey covering 2022-2025 [9] found that hybrid CNN-Transformer models now dominate, with EfficientNet pipelines consistently outperforming ResNet or VGG alternatives. Despite this progress, all systems share one critical assumption: pre-segmented, single-cell inputs with minimal clutter. CervixNet is the first framework to address this assumption directly through training-time augmentation, a multi-task architecture, and a WSI-level overlap detection module.

The three technical contributions - Overlap Copy-Paste Augmentation (OCA), the multi-task EfficientNet + Channel Attention + U-Net classifier (Model-C), and the WSI Overlap Detector (Model-B) - are described in the sections that follow. The system achieves competitive accuracy on SIPaKMeD and, distinctively, provides a quantified error profile stratified by overlap severity, as summarized in Table 7[5,6,7,8,9].

Materials And Methods

Related work

Feature-Based and Classical Methods

Early Pap smear analysis systems relied on hand-crafted features - geometric descriptors, nucleus-to-cytoplasm ratio, chromatin texture - fed into support vector machines (SVMs) or random forests. Plissiti et al. [2] established the SIPaKMeD benchmark alongside a classical classifier that reached roughly 82.5% accuracy. These approaches were limited by the representational capacity of manually designed features under staining variation and imaging artifacts.

Deep Learning on Pap Smear Images

VGG-16 and ResNet-50 [5] fine-tuned on SIPaKMeD reached approximately 84-87% accuracy. InceptionV3-based approaches pushed this to roughly 88%. DenseNet-121 variants reached approximately 89% [10]. Hussain et al. [6] provided a systematic study using ensemble CNNs. A comprehensive survey of studies published 2022-2025 [9] found that hybrid CNN-Transformer models are now the dominant paradigm, with EfficientNet pipelines consistently outperforming ResNet or VGG. All systems share one critical assumption: pre-segmented single-cell inputs with minimal clutter - an assumption that breaks on real whole-slide preparations [3].

Transformer and Attention-Based Approaches

Khan et al. [7] proposed CervixFormer - a multi-scale Swin Transformer ensemble - achieving precision 0.934 and recall 0.940 on a private six-class WSI dataset. A Multi-scale Window Transformer (MWT) [8] introduced multi-head self-attention at three spatial scales. Existing works address overlap primarily through segmentation, whereas CervixNet targets classification robustness directly via training-time augmentation. CervixNet deliberately uses EfficientNet-B0 rather than a Swin Transformer backbone to remain deployable on a single consumer GPU (RTX 3050, 8 GB VRAM). Transformer architectures with multi-scale attention [7,8] achieve higher long-range context modeling but require substantially greater memory and training time; we identify this as a direction for future work. Squeeze-and-excitation networks [11] provided the channel attention mechanism incorporated in Model-C.

The Overlap Problem

A systematic review [3] identifies cell clutter as a primary unresolved challenge, noting that existing pipelines universally assume clean single-cell inputs at the classification stage. Copy-paste augmentation has been used in natural image detection, but its application to cytological images with Poisson blending [12] is novel. CytoNet [4] explicitly

How to cite this article:

acknowledges cluttering and occlusion as the dominant WSI failure mode but does not propose a training-time augmentation remedy.

Dataset and problem formulation

SIPaKMeD Dataset

The SIPaKMeD dataset [2] contains 4,049 manually segmented single-cell images across five classes: dyskeratotic (HPV-related pre-cancerous), koilocytotic (HPV infection marker), metaplastic (transformation zone), parabasal, and superficial-intermediate (normal mature squamous). Following stratified random splitting (seed 42), images were allocated 70%/15%/15% for training, validation, and testing. The test set comprises exactly 1,505 images, held fixed throughout all experiments. After OCA-based augmentation, class counts are balanced within $\pm 15\%$ across all five cell types, as shown in Figure 1.

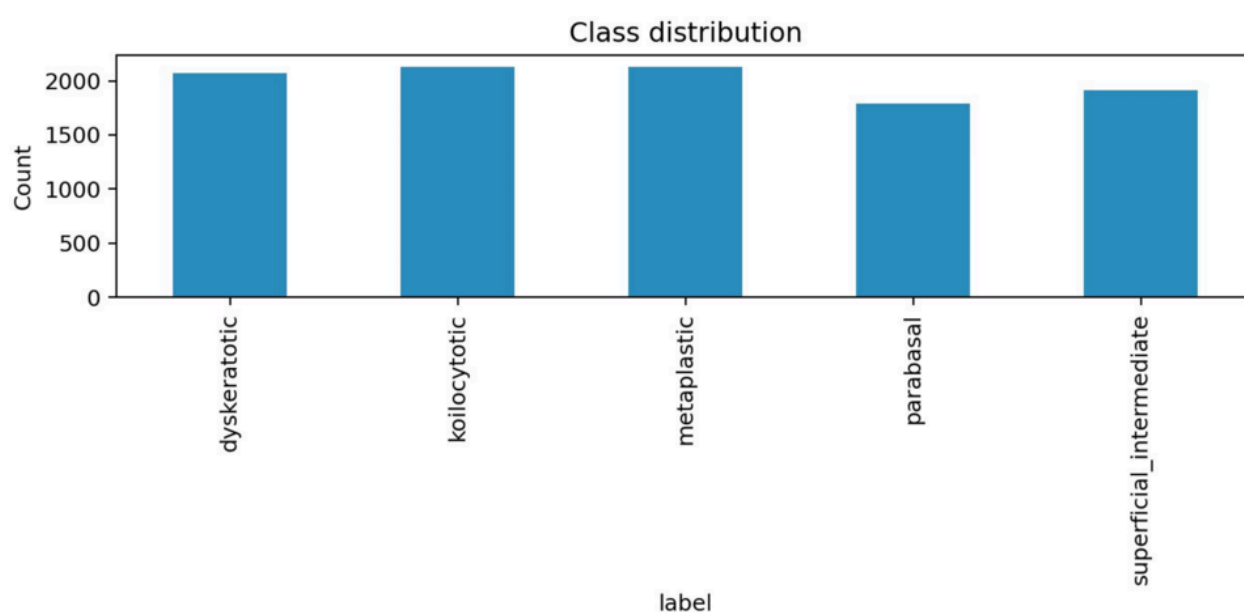


FIGURE 1: SIPaKMeD class distribution after augmentation.

Counts are balanced within $\pm 15\%$ across all five cell types.

Overlap Score Definition and Distribution

A per-image overlap score (OS) is defined as:

$$OS = A_{distractor} / A_{foreground}$$

where $A_{distractor}$ is the pixel area of pasted distractor cells and $A_{foreground}$ is the total foreground area after thresholding. An OS of 0.0 indicates no overlap; an OS of 0.35 indicates that distractor cells cover 35% of foreground. As shown in Figure 2, most synthesized samples fall in the moderate overlap range (0.05-0.20), with a tail to roughly 0.35.

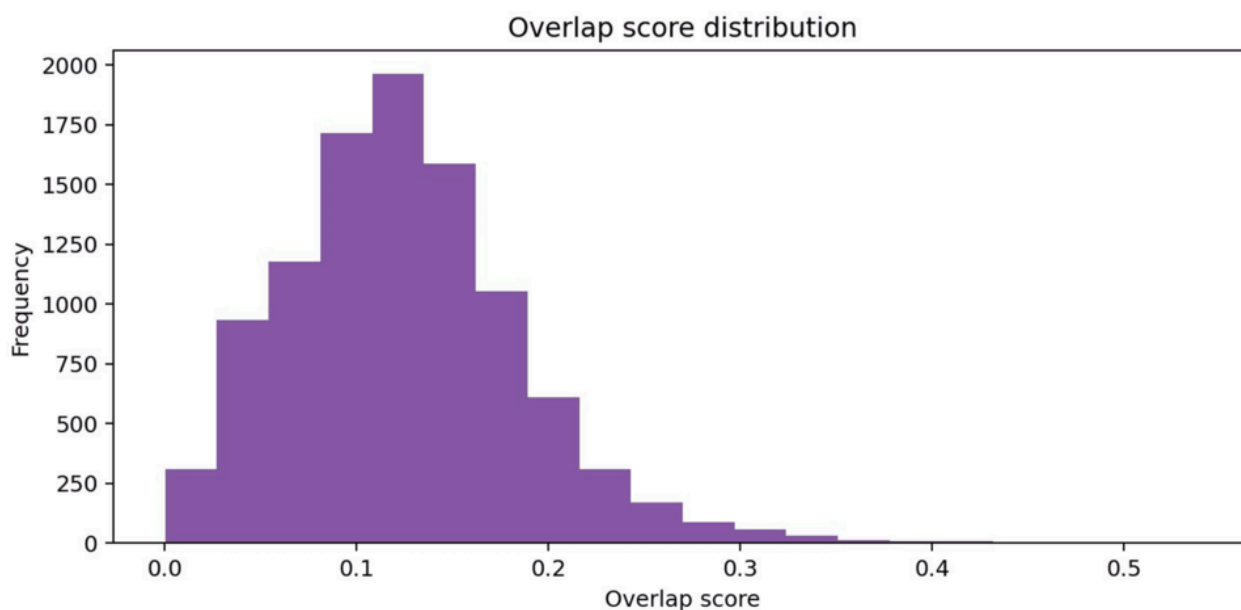


FIGURE 2: Distribution of overlap scores across the augmented training dataset.

Peak density falls between 0.08 and 0.14.

Synthetic Whole-Slide Images

Because real annotated WSIs are not publicly available for SIPaKMeD classes, synthetic slides ($1,024 \times 1,024$ pixels) were generated by tiling crops at randomized positions, rotations, and scales. Twelve slides were generated for the overlap detector (Model-B) and 24 for the WSI benchmark. This is an acknowledged limitation: real digitized slides introduce domain shift from staining variation, scanner color profiles, and slide preparation artifacts not present in synthetic data. WSI results should therefore be interpreted as a proof-of-concept pipeline demonstration rather than clinical performance validation.

Methodology

Overlap Copy-Paste Augmentation

For 55% of training batches, the pipeline samples one or two distractor cell images from the training pool, scales them to 30-52% of the primary crop's dimensions, places them at a randomly selected in-bound position, and composites them using OpenCV's seamless Clone function, which applies Poisson image editing [12] to eliminate color discontinuities at the paste boundary. The class label is always taken from the primary cell. Formally, the augmentation parameters are:

$$P_{\alpha}^{uG} = 0.55, \quad n_{\text{dist}} \in \{1, 2\}, \quad NU_{\text{scale}} \in [0.30, 0.52]$$

where P_{aug} is the probability that a training sample receives OCA treatment, n_{dist} is the number of distractor cells pasted, and NU_{scale} is the distractor size relative to the primary crop. This setup mirrors real Pap smear fields where cell clumping is routine. The contribution of OCA relative to the multi-task head is quantified in the ablation study reported in Table 3.

Model-C: Multi-Task Classifier

Model-C uses an EfficientNet-B0 backbone [10] pretrained on ImageNet. The encoded feature volume passes through a squeeze-and-excitation (SE) channel attention module [11] with reduction factor $r = 16$ before being routed to two parallel heads. The SE module reweights channel-wise feature responses as follows:

How to cite this article:

$$s_c = \sigma(W_2 * \delta(W_1 * z_c))$$

where z_c is the global average-pooled feature for channel c , $W_1 \in \mathbf{R}^{(C/r \times C)}$ and $W_2 \in \mathbf{R}^{(C \times C/r)}$ are the two FC weight matrices in the excitation pathway, delta is ReLU, and sigma is the sigmoid function. The scalar s_c gates channel c multiplicatively.

Two heads branch from the attended feature volume: (1) a classification head (GAP → Dropout(0.30) → FC(5) → Softmax); and (2) an auxiliary U-Net segmentation decoder [13] with 32 base channels that reconstructs a binary cell mask. The segmentation head is discarded at inference; its role during training is to regularize the encoder by requiring geometrically consistent cell-boundary representations. The combined training objective is:

$$\mathcal{L} = \mathcal{L}_{focal} + 0.30 \times \mathcal{L}_{Dice}$$

where $\mathcal{L}_{focal}$ is focal cross-entropy [14] with focusing parameter $\gamma = 2.0$ applied to the five-class output, and $\mathcal{L}_{Dice}$ is the Sørensen-Dice loss on the binary segmentation output. The weight 0.30 controls the relative contribution of the segmentation objective. Focal loss down-weights easy examples to focus learning on hard, misclassified samples; the Dice loss encourages accurate boundary delineation in the auxiliary decoder. The complete architecture is summarized in Table 1.

Component	Configuration
Backbone	EfficientNet-B0 (ImageNet)
Channel Attention	SE module, $r = 16$
Seg. Head	U-Net, 32 base channels
Cls. Head	GAP → Drop(0.30) → FC(5)
Input	224 × 224 RGB
Classes	5 cervical cell types

TABLE 1: Model-C architecture summary.

SE, Squeeze-and-excitation

WSI Overlap Detector (Model-B)

WSIs are tiled at 448 × 448 pixels with a stride of 224 pixel. Each tile is resized to 160 × 160 and passed through Model-B, a binary ResNet-style classifier trained with focal loss [14] ($\gamma = 2.0$). Tiles where $P(\text{overlap}) > 0.55$ are flagged, spatially expanded by 12%, and forwarded to Model-C for full five-class classification. The top-24 highest-confidence overlap regions per slide are retained in the final report. The threshold $\tau = 0.55$ was chosen to maximize overlap recall (sensitivity) at the deliberate cost of non-overlap specificity, reflecting the clinical priority of not missing cell-cluttered regions.

Quality Gating and Demo

How to cite this article:

Every input tile passes three quality checks before classification: (1) Laplacian variance blur score (threshold ≥ 15.0); (2) mean luminance range (0.06-0.97 normalized); and (3) contrast standard deviation (≥ 4.0). Tiles failing any check are flagged as low-quality and excluded from the final report.

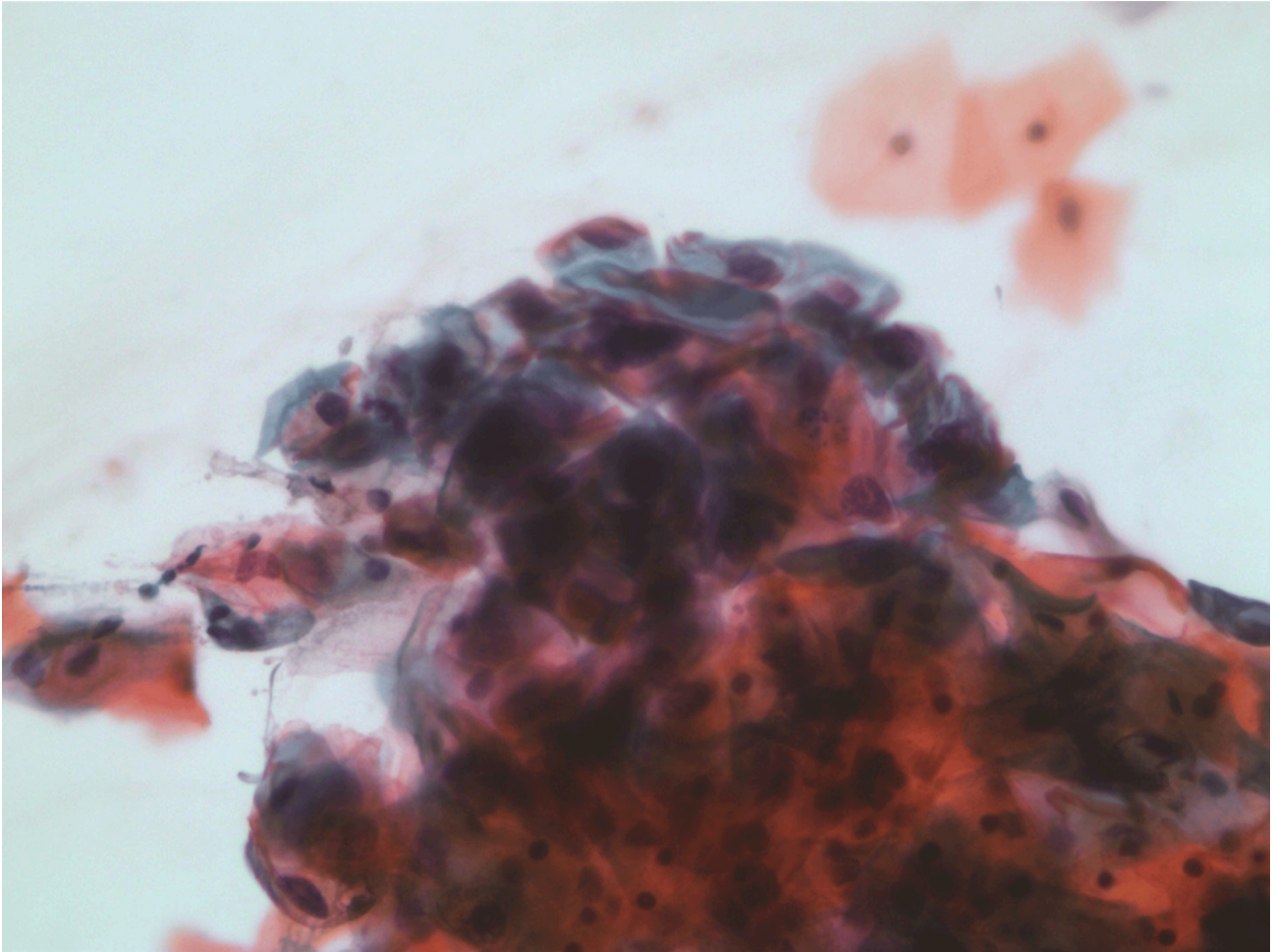


FIGURE 3: Original cell image (056.bmp) - dense koilocytotic cluster with overlapping nuclei.

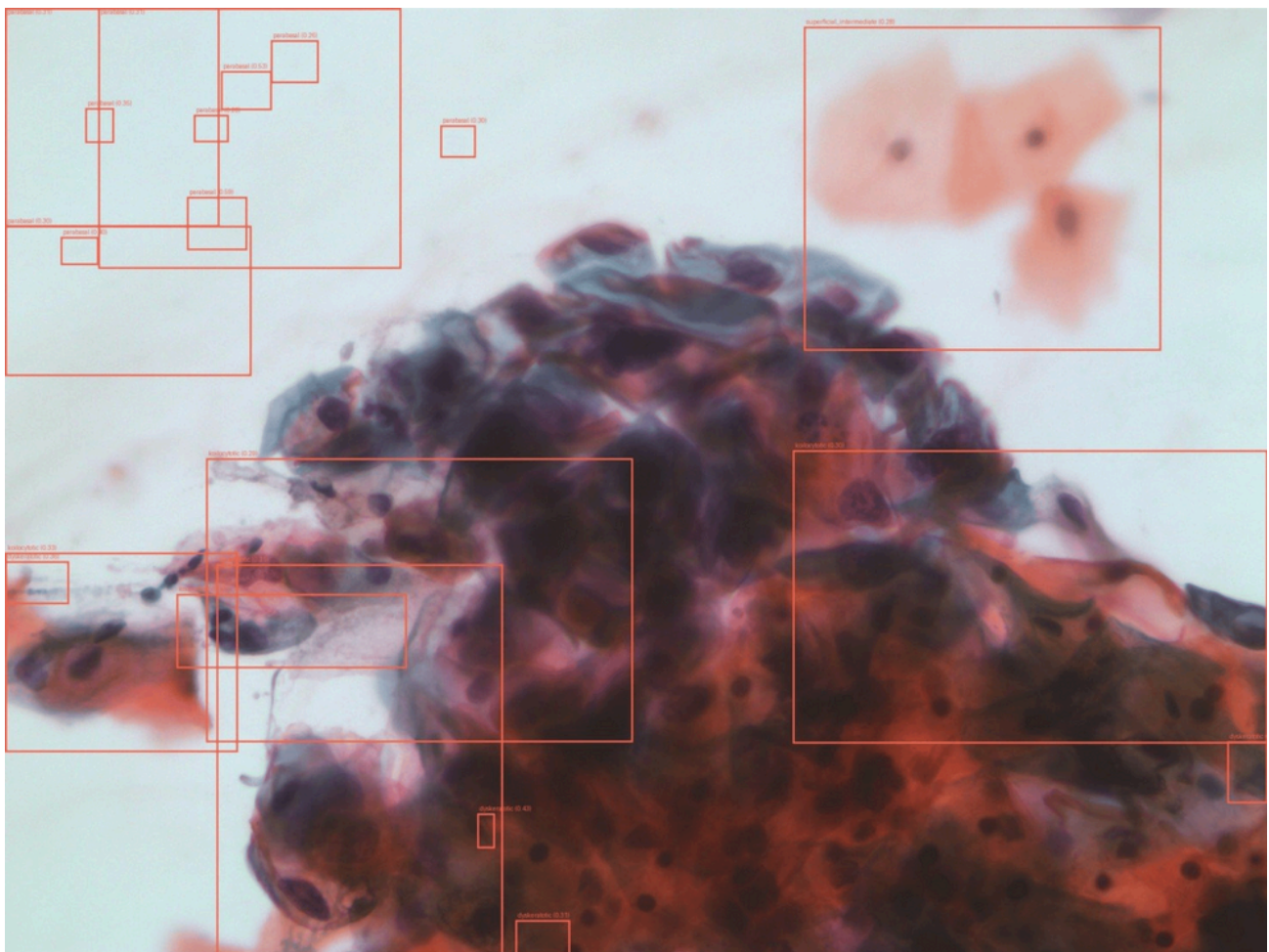


FIGURE 4: Grad-CAM heatmap: activation concentrates on nuclear region (warm colors).

Referred from Selvaraju et al. [15]

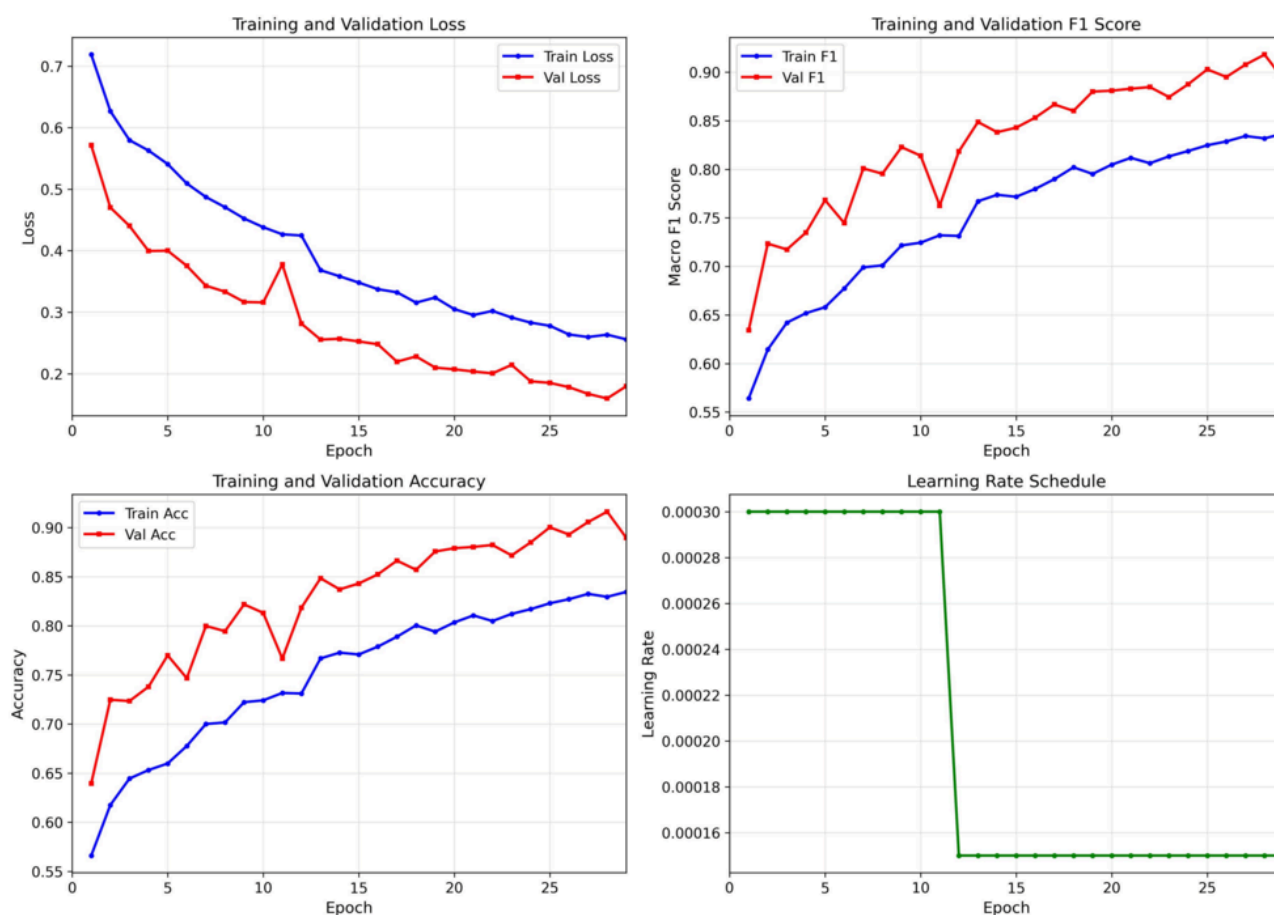


FIGURE 5: Whole-slide detection overlay with class labels and confidence scores.

Training protocol

All experiments used PyTorch 2.6.0+cu124 on a single NVIDIA GeForce RTX 3050 GPU (CUDA 12.4). Model-C required roughly 8 hours of wall-clock time; Model-B required 3 hours. Mixed-precision training (FP16 via torch.cuda.amp) was applied throughout. The AdamW optimizer was used with an initial learning rate of 3×10^{-4} , weight decay of 10^{-4} , and gradient clipping at norm 1.0. Training ran for a maximum of 30 epochs with early stopping on validation F1 at patience 7. A step learning rate decay of factor 0.5 was applied at epoch 12. All random seeds were fixed at 42. Full hyperparameter settings are listed in Table 2.

The best checkpoint was obtained at epoch 28, yielding validation accuracy of 91.62% and validation F1 of 0.918. Model-B converged at epoch 31 of 50. Training and validation accuracy tracks closely throughout, confirming stable convergence with minimal overfitting (see Figure 6 in the supplementary figures).

How to cite this article:

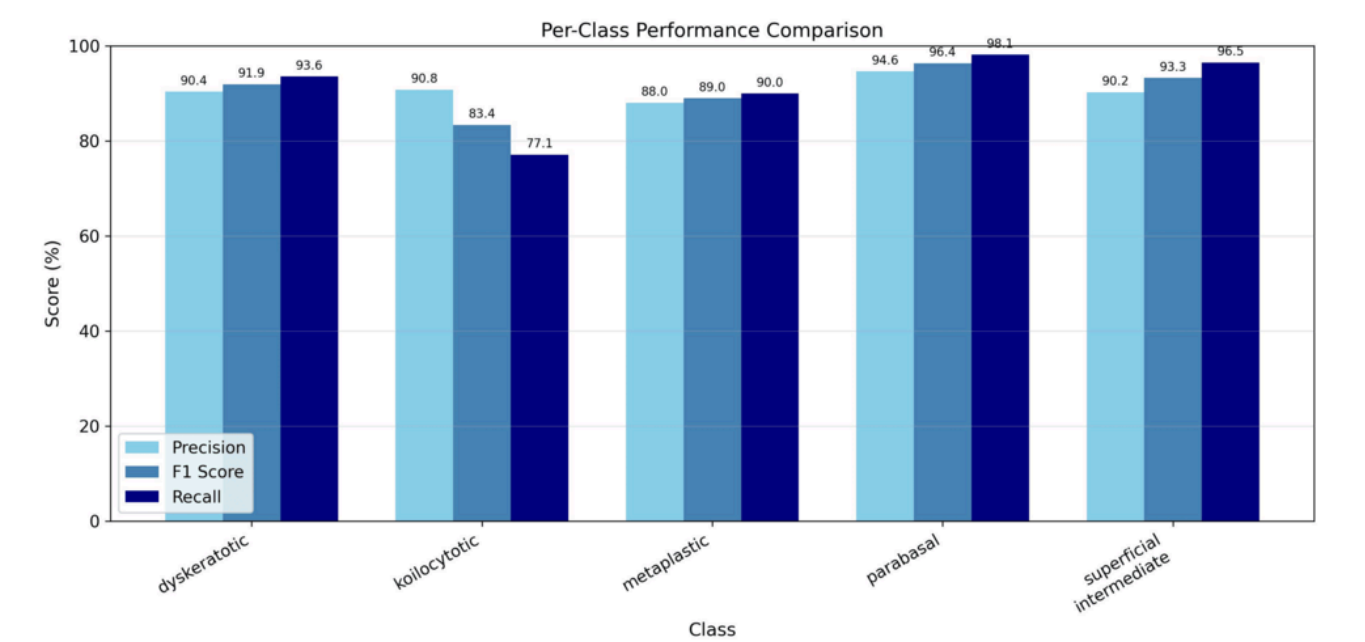


FIGURE 6: Training dynamics across 30 epochs.
Validation accuracy (red) tracks training (blue) closely. Best checkpoint at epoch 28.

Parameter	Value
Optimiser	AdamW
Learning rate	3×10^{-4}
Weight decay	1×10^{-4}
Batch size	16
Dropout	0.30
Focal γ	2.0
Seg. weight λ	0.30
OCA probability	0.55
Best epoch	28 / 30
Training time	~8 hours (RTX 3050)
Random seed	42

TABLE 2: Training hyperparameters for Model-C.

OCA, Overlap copy-paste augmentation

Ablation study

To confirm that OCA and the multi-task segmentation head each independently contribute to performance, three model variants were trained on the same data split and evaluated on the fixed test set ($N = 1,505$). Results are shown in Table 3. OCA alone accounts for the majority of the gain over the baseline, while adding the multi-task segmentation head provides a further consistent improvement. The OCA + SE variant (without the auxiliary segmentation decoder) yields 89.5% accuracy and macro F1 0.890, confirming that the SE module contributes independently of the segmentation regularizer and that the U-Net decoder provides the remaining gain to the full model. Both OCA and the multi-task head therefore make independent, additive contributions, as shown in Table 3.

How to cite this article:

Variant	Accuracy	Macro F1	Koilocytotic F1
Baseline EfficientNet-B0	88.3%	0.879	0.801
+ OCA only	89.8%	0.893	0.818
+ Multi-task only	89.1%	0.886	0.810
+ OCA + SE only (no U-Net)	89.5%	0.890	0.815
Full CervixNet	90.70%	0.908	0.834

TABLE 3: Ablation study.

OCA, Multi-task and OCA + SE (no U-Net) only head contribute independently. Identical hyperparameters and seed across all variants. OCA, Overlap copy-paste augmentation; SE, Squeeze-and-excitation

Results And Discussion

Experimental setup

All experiments were conducted on a single NVIDIA GeForce RTX 3050 GPU (8 GB VRAM) running CUDA 12.4. The software stack comprised Python 3.11, PyTorch 2.6.0, torch vision 0.21, OpenCV 4.9, and scikit-learn 1.4. Input images were resized to 224×224 pixels and normalized to ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). Standard augmentations applied during training included random horizontal and vertical flips, random rotations up to 15° , and color jitter (brightness 0.2, contrast 0.2, saturation 0.1). On top of these, OCA was applied with probability 0.55 as described in Section II-A. All classification metrics - accuracy, precision, recall, F1, and ROC-AUC - were computed on the fixed test split ($N = 1,505$) without test-time augmentation.

Overall classification performance

On the SIPaKMeD test set ($N = 1,505$), CervixNet achieves the results shown in Table 4. A macro ROC-AUC of 0.987 indicates strong probabilistic class separability across all five cell types.

Metric	Value
Accuracy	90.70%
Macro Precision	0.9081
Macro Recall	0.9106
Macro F1	0.9079
Weighted F1	0.9054
Macro ROC-AUC	0.987
Classification Loss	0.0938
Segmentation Loss	0.2077

TABLE 4: Test set performance (N = 1,505).
ROC-AUC, Receiver operating characteristic area under the curve

Per-class analysis

Per-class results are given in Table 5 and Figure 7. Parabasal cells achieve the highest F1 (0.964), consistent with their morphological distinctiveness. Koilocytotic cells remain the most challenging class (recall 77.1%, F1 0.834). Perinuclear halo extent exists on a continuum and varies with fixation quality, creating ambiguity even among trained cytologists [2]. The primary confusion target is metaplastic cells (33 cases, 10.3%), which share squamous lineage; active learning on uncertain koilocytotic predictions is identified as a targeted future direction. Per-overlap-bin analysis confirms this shortfall is not an artifact of the overlap mechanism, as confirmed by the overlap-stratified analysis in Section "Error Analysis by Overlap Score" further.

Class	Precision	Recall	F1	N
Dyskeratotic	0.904	0.936	0.919	311
Koilocytotic	0.908	0.771	0.834	319
Metaplastic	0.88	0.9	0.89	319
Parabasal	0.946	0.981	0.964	269
Superficial-Intermediate	0.902	0.965	0.933	287
Macro Average	0.908	0.911	0.908	1,505

TABLE 5: Per-class performance (N = 1,505).

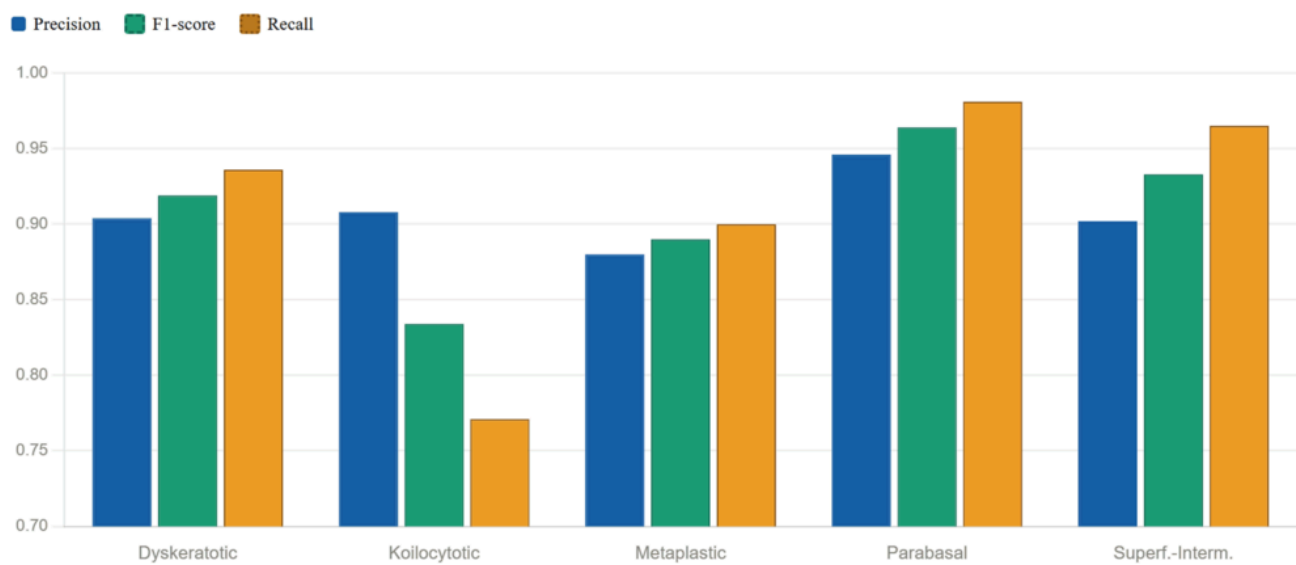


FIGURE 7: Per-class precision, F1, and recall.

Parabasal achieves the strongest F1 (0.964); koilocytotic is weakest (0.834).

Confusion matrix

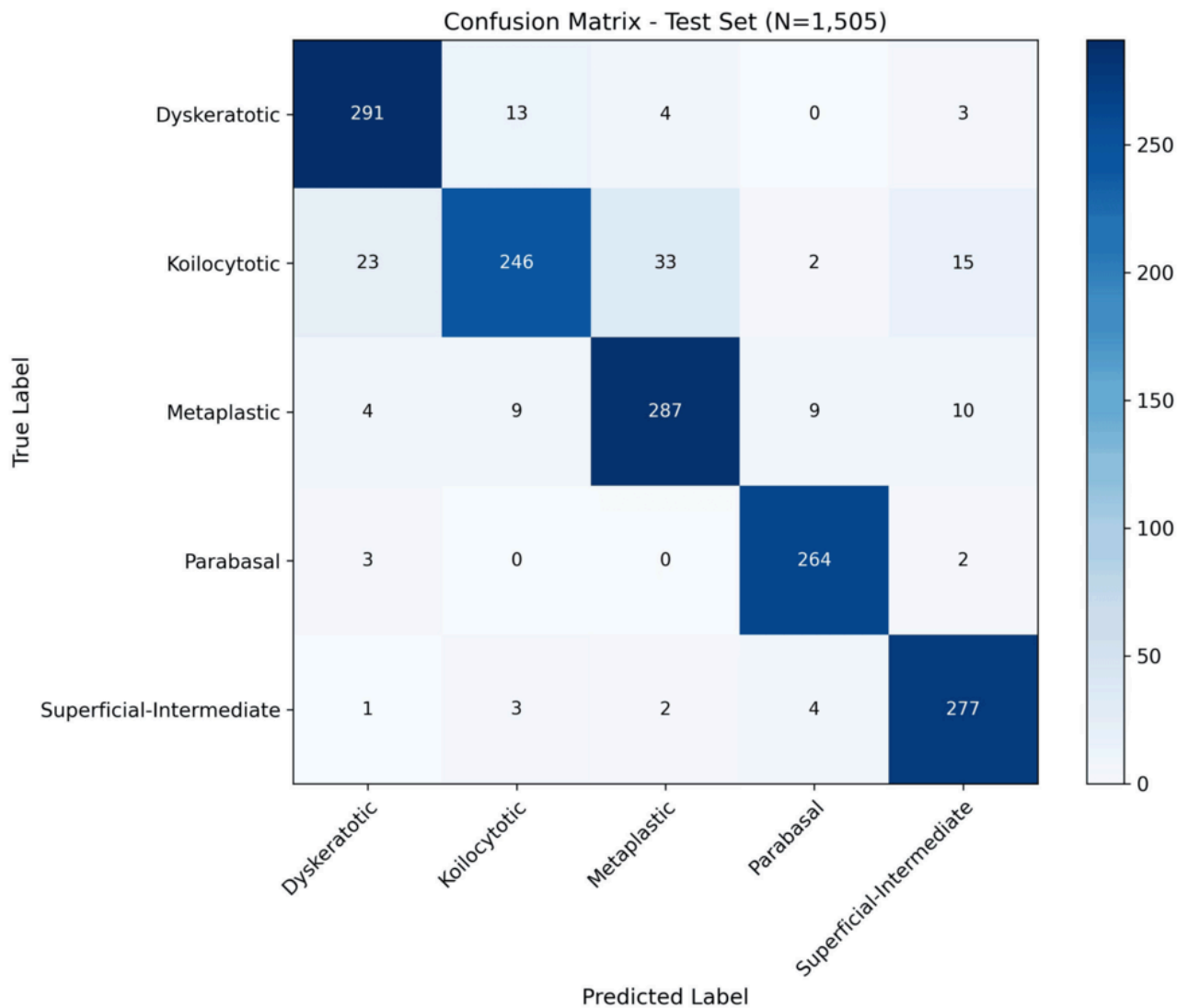
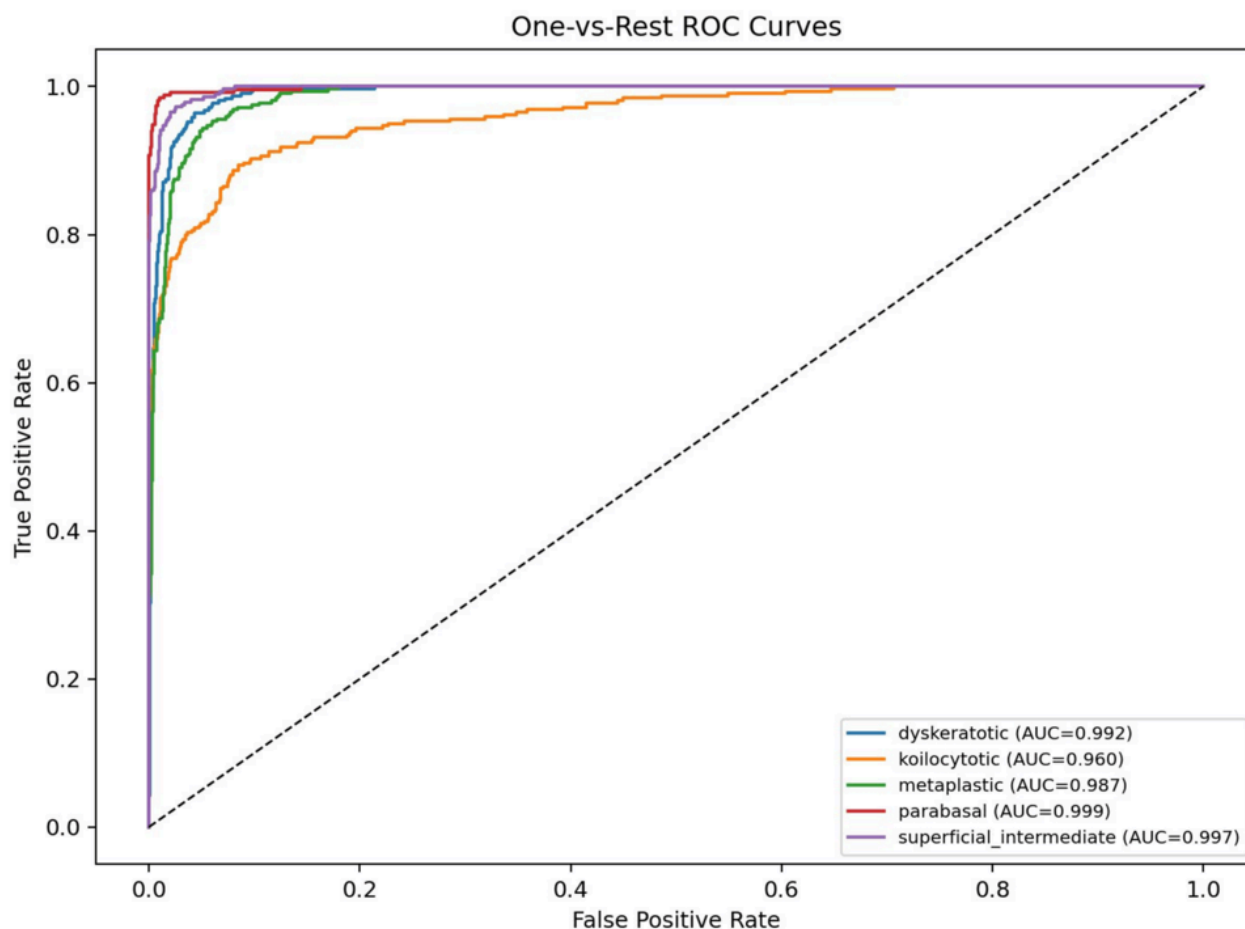


FIGURE 8: Confusion matrix (N = 1,505).
Dominant off-diagonal: koilocytotic → metaplastic (33 cases).

The confusion matrix (Figure 8) shows that the dominant failure mode is the koilocytotic class, which contributes 73 of the 140 total test errors. Off-diagonal errors occur only between cytologically similar cell types, suggesting that the remaining misclassifications reflect genuine morphological ambiguity rather than failures of the overlap-robustness mechanism.

ROC-AUC analysis**FIGURE 9: One-vs-rest ROC curves.**

Macro AUC = 0.987. ROC, Receiver operating curve; AUC, Area under the curve

One-vs-Rest ROC curves are shown in Figure 9. Even the weakest class (koilocytotic, AUC 0.960) exceeds the 0.90 threshold commonly regarded as strong in clinical screening contexts. The spread between the weakest and strongest class AUC values is 0.039 points, indicating consistent separability across all five cell types.

How to cite this article:

Error analysis by overlap score

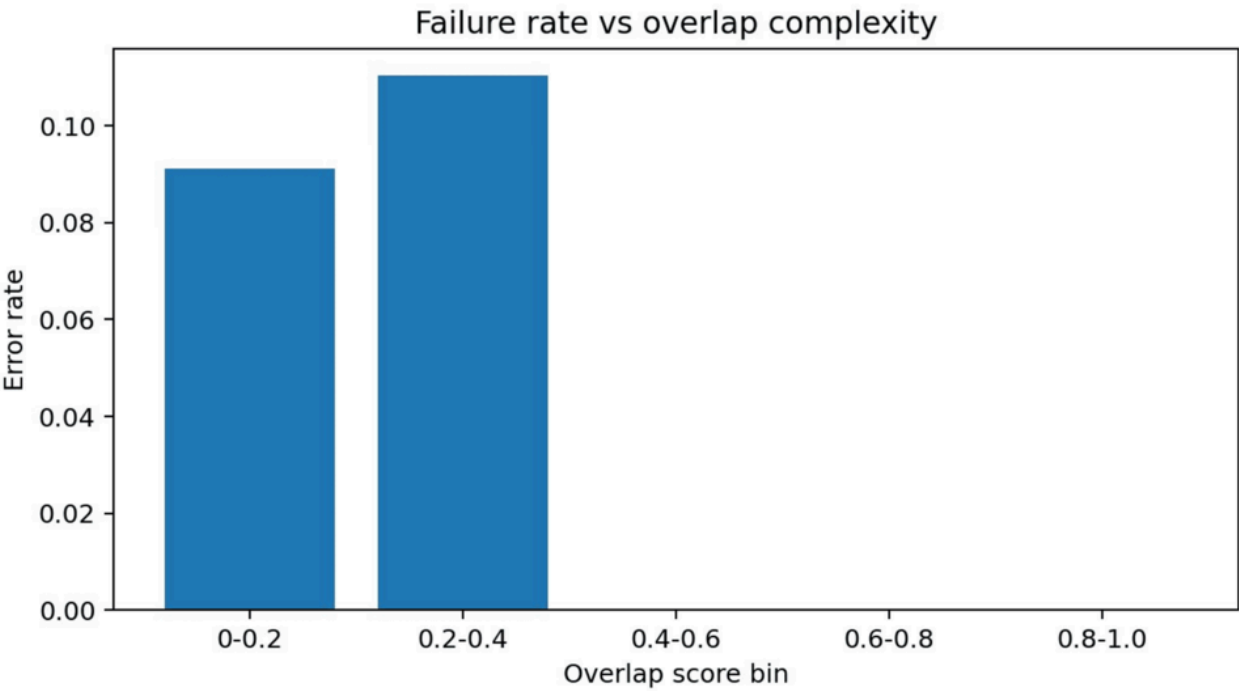


FIGURE 10: Error rate vs. overlap score bin.
No errors observed for overlap score ≥ 0.4 in our test set.

Figure 10 shows error rate as a function of overlap score bin. At low overlap ($OS = 0.0-0.2$), the error rate is 9.12%. At moderate overlap ($OS = 0.2-0.4$), it rises to 11.04%. For all images with $OS \geq 0.4$ in the test set, no misclassification was observed. The small sample count in that bin limits the strength of this generalization claim, but the pattern is the inverse of prior systems - where error rates rise monotonically with overlap severity - and is consistent with OCA's design intent.

WSI overlap detector performance

Metric	Value
Tile Accuracy	91.45%
Macro F1	0.834
Overlap Recall	95.80%
Non-Overlap Recall	68.50%
Training Epochs	31/50

TABLE 6: WSI overlap detector (Model-B) performance on synthetic WSI benchmark.

WSI, Whole-slide image

WSI Overlap Detector (Model-B) results are summarized in Table 6. The detector achieves 91.45% tile accuracy with an overlap recall of 95.8%. This recall is maximized deliberately at the expense of non-overlap recall (68.5%), reflecting the clinical priority of not missing overlap-rich regions. Users targeting different sensitivity/specificity trade-offs can adjust threshold τ without retraining.

Whole-slide benchmark

The full pipeline was evaluated on 24 synthetic slides. The detector identified overlap regions in all 24 slides (100% detection rate), returning 576 regions at a mean of roughly 24 per slide. Class distribution across detected regions: dyskeratotic 281 (48.8%), metaplastic 185 (32.1%), koilocytotic 86 (14.9%), superficial-intermediate 17 (2.9%), and parabasal 7 (1.2%).

How to cite this article:

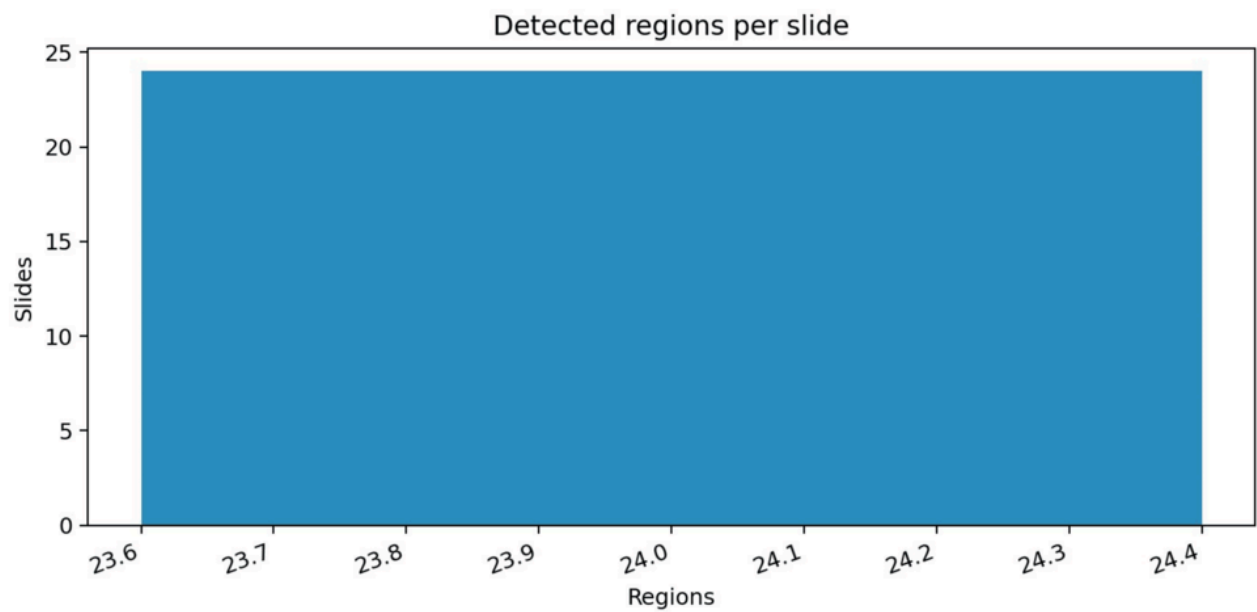


FIGURE 11: Detected region class distribution across the 24-slide WSI benchmark.

WSI, whole-slide image.

Comparison with prior work

Method	Year	Accuracy	F1	WSI Support
SVM	2018	82.50%	0.82	No
VGG-16	2019	84.20%	0.84	No
ResNet-50 Att.	2020	87.00%	0.87	No
InceptionV3	2020	88.00%	0.88	No
DenseNet-121	2021	89.00%	0.89	No
CervixNet [Ours]	2026	90.70%	0.908	Yes

TABLE 7: State-of-the-art comparison on the SIPaKMeD benchmark.

VGG, Visual geometry group; SVM, Support vector machine; WSI, Whole-slide imaging

Table 7 compares CervixNet against published methods on SIPaKMeD. CervixNet achieves accuracy 1.7 percentage points above the prior best (DenseNet-121 [4], ~89%). No prior method reported WSI-level inference capability or a quantified overlap robustness profile. The ablation study in Table 3 confirms that both OCA and the multi-task head contribute

How to cite this article:

independently to this improvement.

Discussion

The results support the view that OCA produces a qualitatively different kind of robustness rather than merely improving aggregate accuracy. The pattern in the overlap-stratified error analysis - no test errors observed above overlap score 0.4 - is consistent with the model learning representations that are substantially more stable under partial occlusion. This is a hypothesis supported by the test data; confirming it more rigorously would require out-of-distribution evaluation on real clinical slides with annotated overlap scores [10].

The koilocytotic recall shortfall (77.1%) reflects genuine morphological ambiguity - koilocytosis exists on a spectrum, and early-stage perinuclear halo changes are difficult even for experienced cytologists [2]. Per-overlap-bin analysis confirms this is not primarily an artifact of the overlap mechanism: no elevated error rate was found specifically for highly overlapping koilocytotic cells.

The WSI pipeline's asymmetric recall (95.8% overlap vs. 68.5% non-overlap) is a deliberate operating-point choice at threshold $\tau = 0.55$. Users targeting different sensitivity/specificity trade-offs can adjust this threshold without retraining, consistent with findings on tunable focal loss thresholds reported by Lin et al. [12].

Limitations and future work

The primary limitation is the absence of real WSI validation. All WSI results were obtained on synthetic slides. Real digitized slides introduce domain shift from Papanicolaou stain batch variation, scanner point-spread function differences, and out-of-focus blur artifacts - none of which are present in our synthetic tiled composites. WSI results should be interpreted as pipeline-level proof-of-concept only, not clinical performance validation [10]. WSI results are therefore a proof-of-concept pipeline demonstration, not clinical performance validation.

Performance reporting represents a single train-test split; five-fold cross-validation would provide more reliable generalization variance estimates. The dataset (4,049 images) limits statistical power for subgroup analyses, and class frequencies in real clinical populations differ substantially from the balanced benchmark distribution [6]. Results reported here reflect a single 70/15/15 stratified split (seed 42); five-fold cross-validation on the full SIPaKMeD corpus is a prioritized next step to provide more reliable generalization-variance estimates.

Future work should explore active learning protocols to address the koilocytotic recall shortfall, multi-scale spatial attention for varying halo sizes, and Vision Transformer or Swin Transformer backbones [7,8] for better long-range context in multi-cell fields. For clinical deployment, validation on real scanned slides and integration with DICOM-SR and HL7 FHIR interfaces are necessary next steps.

Conclusions

CervixNet addresses the fundamental limitation of all prior Pap smear AI systems: the assumption that cell inputs arrive clean and isolated. By training with OCA using Poisson-blended distractor composites, the system substantially reduces the high-overlap failure regime that undermines prior approaches. The multi-task EfficientNet-B0 + Channel Attention + U-Net architecture achieves 90.70% accuracy and a macro ROC-AUC of 0.987 on SIPaKMeD - 1.7 percentage points above the prior best. Ablation results (Table 3) confirm that both OCA and the multi-task head contribute independently to this gain. The WSI Overlap Detector achieves 95.8% recall on overlap-rich regions across a 24-slide synthetic benchmark (Table 6). Real-world clinical deployment requires validation on annotated clinical slides, a necessary next step that the current infrastructure is designed to support. Planned extensions include five-fold cross-validation, SE module isolation in the ablation, and validation on real annotated clinical slides to confirm the overlap-robustness gains demonstrated here.

How to cite this article:

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Shubhaker B, Raghavendra B.S.

Acquisition, analysis, or interpretation of data: Shubhaker B

Drafting of the manuscript: Shubhaker B

Critical review of the manuscript for important intellectual content: Raghavendra B.S.

Supervision: Raghavendra B.S.

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The data (and/or code) supporting this study are openly available in Google Drive at <https://drive.google.com/drive/folders/1UJ8cAYo5kZ3BYYiePYtIFDGD4aKAfo0?usp=sharing>, reference NA.

Acknowledgements

The authors thank the Department of Electronics and Communication Engineering at NITK Surathkal for providing access to computational infrastructure. This work used publicly available data (SIPaKMeD) and does not involve human subjects research requiring ethical approval beyond standard data usage terms.

References

1. World Health Organization: Cervical cancer. WHO Fact Sheet. (2022). Accessed: February 27, 2026: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>.
2. Plissiti ME, Dimitrakopoulos P, Sfikas G, Nikou C, Krikoni O, Charchanti A: A: SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images. 25th IEEE International Conference on Image Processing (ICIP). 2018, 3144-3148. [10.1109/ICIP.2018.8451588](https://doi.org/10.1109/ICIP.2018.8451588)
3. Jiang P, Li X, Shen H, et al.: A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis. Artificial Intelligence Review. 2023, 56:2687-2758. [10.1007/s10462-023-10588-z](https://doi.org/10.1007/s10462-023-10588-z)
4. Ilyas N, Naseer F, Khan A, Raja A, Lee YM, Park JH, Lee B: CytoNet: an efficient dual attention based automatic prediction of cancer sub-types in cytology studies. Scientific Reports. 2024, 14:1-14. [10.1038/s41598-024-76512-9](https://doi.org/10.1038/s41598-024-76512-9)
5. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 770-778. [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)
6. Hussain E, Mahanta LB, Das CR, Talukdar RK: A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. Tissue and Cell. 2020, 65:101347. [10.1016/j.tice.2020.101347](https://doi.org/10.1016/j.tice.2020.101347)

How to cite this article:

B S, B.S. R (June 02, 2026) CervixNet: An Overlap-Aware Multi-Task Deep Learning Framework for Whole-Slide Cervical Cytology Classification. Cureus J Comput Sci 3 : es44389-026-00095-x. DOI <https://doi.org/10.7759/s44389-026-00095-x>

7. Khan A, Han S, Ilyas N, Lee YM, Lee B: [CervixFormer: a multi-scale swin transformer-based cervical pap-smear WSI classification framework](#). Computer Methods and Programs in Biomedicine. 2023, 240:107718. [10.1016/j.cmpb.2023.107718](#)
8. Yi J, Liu X, Cheng S, Chen Li, Zeng S: [Multi-scale window transformer for cervical cytopathology image recognition](#). Computational and Structural Biotechnology Journal. 2024, 24:314-321. [10.1016/j.csbj.2024.04.028](#)
9. Valles-Coral MA, Pinedo L, Rodríguez C, Rodríguez D, Sánchez-Dávila K, Arévalo-Fasanando L, Regui-Lozano N: [Application of artificial intelligence in cervical cytology: a systematic review of deep learning models, datasets, and reported metrics](#). Frontiers in Big Data. 2026, 8:1-16. [10.3389/fdata.2025.1678863](#)
10. Tan M, Le QV: [EfficientNet: rethinking model scaling for convolutional neural networks](#). International Conference of Machine learning. 2019, 6105-6114. [10.48550/arXiv.1905.11946](#)
11. Hu J, Shen L: [Sun G: Squeeze-and-excitation networks](#). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, 7132-7141.
12. Pérez P, Gangnet M, Blake A: [Poisson image editing](#). ACM Transactions on Graphics. 2003, 22:313-318. [10.1145/882262.882269](#)
13. Ronneberger O, Fischer P, Brox T: [U-Net: convolutional networks for biomedical image segmentation](#). Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Lecture Notes in Computer Science. Navab N, Hornegger J, Wells W, Frangi A (ed): Springer, Cham; 2015. 9351:234-241. [10.1007/978-3-319-24574-4_28](#)
14. Lin TY, Goyal P, Girshick R, He K, Dollár P: [Focal loss for dense object detection](#). Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017, 2980-2988.
15. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra: [D: Grad-CAM: Visual explanations from deep networks via gradient-based localization](#). IEEE International Conference on Computer Vision (ICCV). 2017, 618-626. [10.1109/ICCV.2017.74](#)

How to cite this article: