Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FASTER YOLO: AN EFFICIENT FRAMEWORK FOR CERVICAL CANCER CELL DETECTION WITH DEFORMABLE CONVOLUTIONAL ATTENTION

JHEELAM MONDAL[1], RAJDEEP CHATTERJEE[1] , MAHENDRA KUMAR GOURISARIA[1,*]
[1]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar-751027, Odisha, India
E-mail: [1]mkgourisaria2010@gmail.com

ABSTRACT

Cervical cancer is the fourth most common disease worldwide. The most common diagnostic method required for cervical cancer screening is the pap smear test. Making precise diagnosis, identifying and classifying cells and closely examining each slide all take a significant amount of time and work. Long stretches of visual inspection can make human mistakes more likely thereby resulting in incorrect classification of cells. An essential stage in automatic cytopathology diagnosis is the detection of nuclei in cervical cell images. In recent years, YOLO (You Only Look Once) models have been the most popular paradigm in the field of real-time object detection because of their successful balance between detection performance and processing cost. This work focuses on a number of YOLO models and various cutting-edge object detection methods that are trained on the popular SIPAKMED benchmark dataset. This dataset contains annotated labels for each image. In this paper, we provide an improved YOLO-based object detection model that achieves performance comparable to state-of-the-art YOLO models while dramatically decreasing computing complexity. The proposed model Faster YOLOv13s is built with an optimized attention aware architecture that prioritizes efficiency above detection accuracy. Experimental results show that the proposed model achieved a competitive mAP_{50} score of 87.00% compared to the best-performing YOLO model while significantly reducing the number of trainable parameters and taking significantly less training time. The results of this study are meant to guide future clinical applications and identify the best model for cervical cancer detection.

Keywords: *Cervical Dysplasia, Image Pre-Processing, Object Detection, YOLO Network, Deep Neural Network*

1. INTRODUCTION

Cervical cancer is the fourth most prevalent disease in the world in Globocan 2022, with 3,50,000 deaths and 6,60,000 new cases. The prevalence of cancer has remained elevated globally in recent years. 94% of cervical cancer fatalities in 2022 occurred in low and middle-income nations. In nations with low and moderate incomes, women are at a greater risk for cervical cancer because of a lack of awareness and limited access to health services [1]. Numerous nations are collaborating in an effort to eradicate cervical cancer by 2030. According to a recent study by Johns Hopkins researchers, false positives occur in between one and ten percent of Pap tests [2]. Since there are no early signs of cervical cancer, the routine pap smear test is the primary screening approach for the disease's early detection and prevention [3]. The pap smear test is one of the most widely used screening tools for detecting cervical cancer cells or malignant cells, depending on the detailed microscopic findings. For this test, a glass slide with a sample of cells that have been collected is stained using a mixture of stains called the Papanicolaou stain. A pathologist then examines the slide to look for any anomalies in the cells. Since there are 100-10,000 cells in a pap-smear image, it is a difficult as well as time-consuming process for a pathologist to carefully examine each one under a microscope. In this work, we have used the various YOLO models to identify the aberrant cervical cells, guaranteeing the detection's speed and precision [4-5]. Many object detection algorithms are based on conventional Convolutional Neural Networks (CNNs), but the YOLO technique has a number of important advantages, especially when it comes to speed and real- time performance. Several steps are frequently involved in conventional CNN-based object detection techniques. First one is selection of a region which is a possible area of interest based on an algorithm. Next the CNN

classifier extract features from the suggested regions and identifies the object class. Lastly the overlapping bounding boxes are eliminated using Non-Maximum Suppression (NMS) algorithm. Two-stage detectors include Faster Region-based Convolutional Neural Network (Faster R-CNN) and Region-based Convolutional Neural Network (R-CNN) [6]. In contrast, object detection is handled by YOLO algorithm as a single regression problem. In a single forward pass, the neural network creates a grid out of the image and predicts bounding boxes and class probabilities for each grid cell. Real-time object detection is made possible by this unified design, which also greatly speeds up YOLO models and lowers computing overhead. Its exceptional performance is largely due to its single-pass architecture, end-to-end training and global contextual comprehension [7].

Our work includes variants of YOLO models such as YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, YOLOv12, YOLOv13 and different state-of-the-art techniques such as Faster R-CNN and Roboflow Detection Transformer (RF-DETR). Using a CNN for object detection was revolutionized by Faster R-CNN [8]. Before this, most architectures used hand-crafted features for object detection. This architecture is a two-stage detector model, that suggests a potential region where objects might be present and then predict the bounding boxes and classes. This two-step process makes it significantly slower than the single-stage models like YOLO, which predict the object bounding boxes and classes in one go. RF-DETR is a transformer-based real time model [9]. Although it aims to bridge the gap between accuracy and speed, this model architecture may be more intricate than the CNN-based YOLO models. Our experimental setup is applied to nano, small and medium variants of all the mentioned YOLO models, and rigorously all the models are trained using the complex cervical cancer SIPAKMED dataset. The accuracy of the YOLOv13s model has been found to be significantly higher than other SOTA techniques using the same setup. The paper is sub-divided into five sections. The following are included in later sections of the paper: Some of the recent related works is covered in the second section, “Literature Review”, the third section contains the “Materials and Methods”, the fourth section consists of “Results and Discussion”, the fifth section covers the “Limitations” and the sixth section discussed the ‘Conclusion and Future work’ followed by References.

2. LITERATURE REVIEW

In recent years, researchers around the world have conducted numerous studies to identify and predict cervical cancers using various methodologies. On the liquid-based cytology dataset, the transfer learning of cervical exfoliated cells using Faster R-CNN was accomplished using the weight parameters derived from the ImageNet dataset pretraining. Xueyu *et al.* 2019 [10] conducted the experiments using ResNet-101 as the basis to learn the depth and superficial aspects of cervical exfoliated cells and have achieved a mAP value of 66.98%. The primary goal of this work is to use multi-semantic labels to divide overlapping regions into several single cells. The findings show that the overlapped cells may be effectively separated into numerous single cells, although the classification effect varies, with a low probability of 54.00% and a high probability of 95.00%. To achieve more reliable detection and classification results for overlapping regions, additional research must be done on the integration of region-based morphological recognition and semantic recognition. Xiang *et al.* 2020 [11] in their work have used YOLOv3, one of the most advanced CNN-based object detection techniques models as the baseline model and applied on a private liquid-based cytology dataset. They have cascaded an extra task-specific classifier to enhance the classification performance of the four class examples, which are very similar categories. In this work, in addition to classifying cervical cells at the image level, this automation-assisted cervical cell reading method offers more precise location and category reference data for abnormal cells. This model obtained the best mAP value of 63.40%.

Sampaio *et al.* 2021 [12] achieved satisfactory results in the localization and recognition of various types of cervical cells from conventional cytology images using the faster R-CNN approach. This is the first investigation conducted on the public SIPAKMED dataset, achieving 33.70% mAP value. Further work can be investigated on hybrid pipelines with intermediate modules that separate the diverse cell types that define related lesion classes. At the same time, the dataset has to be significantly enhanced in terms of both class representation and data volume. Additionally, it is important to overcome the subjectivity of the annotation process by focusing on standardizing the recognized objects of interest. Using the SIPAKMED dataset, the study by Lohith *et al.* 2023 [13] suggests an integrated convolutional neural network (CNN)-based technique for the detection and classification of cervical cell

abnormalities. It can correctly identify the class of several cell types found in the multi-cell test sample. The YOLOv5l model achieved an average mAP score of 59.00%.

Kalbhor *et al*. 2023 [14] have implemented three object detection models namely, Detectron2, developed by Facebook AI Research (FAIR) Group; Faster R-CNN, that incorporates Region Proposal Network (RPN) and YOLOv5, which employs the CSPNet backbone. The Roboflow tool is used to pre-process and enhance the CRIC dataset. The performance measures of average precision and means average precision over the Intersection over Union (IoU) are employed to assess the efficacy of the model. The YOLOv5 model for binary classification has achieved the highest mean Average Precision (mAP) of 82.40% on the augmented dataset. The diversity and unpredictability of pap smear images in the actual world might not be well represented by the CRIC dataset used to train the models. A small dataset

Table 1: Analysis of some research articles

Author/Year	Method	Dataset Used	mAP	Limitations
Xueyu *et al.* 2019 [10]	Faster R-CNN (ResNet-101 backbone)	Private LBC Dataset (6 classes)	66.98	Integration of semantic recognition with region-based morphological recognition is required to provide more reliable detection and classification results for overlapping cells.
Xiang *et al.* 2020 [11]	Custom YOLOv3	Private LBC dataset	63.40	The results show that the automatic detection method's performance serves as a solid foundation and point of reference for the upcoming work.
Sampaio *et al.* 2021 [12]	Faster R-CNN	SIPAKMED	33.70	A more cost-effective mobile IoT framework can be developed in future by keeping the analysis time per cytological sample less than 4 minutes. Its incorporation into a computer-aided diagnosis system requires further enhancements.
Lohith *et al.* 2023 [13]	YOLOv5l	SIPAKMED	59.00	The work is limited to basic CNN for detection and classification work. Strong object detection frameworks such as Faster R-CNN with better back-bones/FPN known for high accuracy can be explored in future.
Kalbhor *et al.* 2023 [14]	YOLOv5	CRIC dataset (binary class)	82.40	With more distinct samples, a multi-class cancer cell prediction model can be developed that can both match and surpass the predictions of binary classification between normal and abnormal.
Ontor *et al.* 2023 [15]	YOLOv5m	Intel Mobile ODT dataset	86.60	More sophisticated technologies will be used in future to enhance the suggested system with additional images in order to boost the model's effectiveness.
Wu *et al.* 2024 [16]	DSIR-YOLO (based on YOLOv5)	Private LBC dataset	70.50	DSIR-YOLO model has the inability to inability to identify small-sized cell targets in dual-stained pathological images.
Indugu *et al.* 2025 [17]	CerviYOLO (based on YOLOv5)	Herlev (7-class)	95.40	Though it performs well on this dataset, other datasets can be explored to check the robustness of the proposed model.

might not account for all potential differences in cervical cells and could result in restricted generalization. Conventional pap smear samples are included in the CRIC dataset, and since the models are general, they can be used with liquid-based cytology (LBC) samples. Additionally, there is a lot of image quality distortion in pap smear images. To increase model's accuracy, image quality can be improved without losing morphological information. For cervical cancer screening to be automated and early diagnosis, accurate cell segmentation is essential. Ontor *et al*. 2023 [15] have implemented four variants of YOLOv5 model to detect the cervical cancerous cells. YOLOv5m beat all YOLOv5 model versions, even if YOLOv5s performed similarly to YOLOv5m. Using the cervix images, this low-cost and automated method will help patients and physicians identify malignant cells. The YOLOv5m model achieved 88.40% precision, 86.40% recall, and achieved the highest mAP value of 86.60%. This model also gained the least object loss, box loss and classification loss values.

A novel method with good sensitivity and specificity for cervical cancer screening is the p16/Ki-67 dual staining technique. However, when YOLOv5s method is directly applied to dual-stained cell images, there are problems with mis-detection and incorrect recognition. Wu *et al*. 2024 [16] in their research, has proposed a novel dual-stained image recognition model for cervical cancer (DSIR-YOLO) based on a YOLOv5 model. The detection performance is greatly enhanced by combining the GAM attention mechanism, multi-scale feature fusion, EIoU loss function and Swin-Transformer module; mAP@0.5 and mAP@0.5:0.95 achieve 92.6% and 70.5% respectively. In five-fold cross-validation, the updated method outperforms YOLOv5s by 2.3%, 4.1%, 4.3% and 8.00% in accuracy, recall, mAP@0.5 and mAP@0.5:0.95. Indugu *et al*. 2025 [17] has suggested CerviYOLO model that shows great promise for enhancing cervical cancer detection accuracy and efficacy by utilizing YOLOv5's improved capabilities. This model achieves a mAP value of 95.40% and an F1-score of 92.70% by successfully classifying seven different classes of cervical cells present in Herlev dataset, demonstrating its resilience in detecting cervical abnormalities. The more detailed analysis on the literature survey is again depicted in the above Table 1.

3. MATERIALS AND METHODS

This section covers the materials and methodology used to classify cervical cancer images.

3.1 Dataset Description

The SIPAKMED dataset [18], which is accessible to the public, is used to conduct experimental investigations. This largest publicly available benchmark dataset consists of Pap smear slides with 4049 isolated cell images out of 966 cluster cell images. Based on the cells' form and appearance, skilled cytopathologists have classified them into five groups. Below Table 2 displays the SIPAKMED distribution of cervical cell images for each class.

Table 2: Distribution of Image class labels from SIPAKMED dataset

Class Name	Cell Count
Superficial-Intermediate	813
Parabasal	787
Koilocytotic	825
Metaplastic	793
Dyskeratotic	813
Total	4049

The Roboflow Universe is a big, open-source collection of computer vision datasets and trained models. One major benefit is that a large number of datasets are either readily exportable to or already in YOLO format. This simplifies the data preparation procedure, which frequently acts as a bottleneck in projects involving object detection task. We have utilized this feature to download the annotated data from Roboflow Universe. Following Table 3 shows SIPAKMED dataset and the annotations present across training, testing and validation samples. A train/validation/test split of 87%, 9% and 4% respectively is a popular distribution used in Roboflow annotation datasets applied on YOLO models. In Table 3, Class0, Class1, Class2, Class3 and Class4 corresponds to Dyskeratotic, Koilocytotic, Metaplastic, Parabasal and Superficial-Intermediate classes respectively.

Table 3: Distribution Of Image Class Labels From SIPAKMED Dataset

Name	Number of Images	Number of Objects	Class0 objects	Class1 objects	Class2 objects	Class3 objects	Class4 objects
Train Dataset	3389	13986	3133	3272	2985	2963	1633
Valid Dataset	348	1343	268	285	315	280	195
Test Dataset	169	701	189	147	132	135	98
Total Dataset	3906	16,030	3590	3704	3432	3378	1926

3.2 Dataset Visualization

Following Figure 1 shows the sample images from this dataset. Also, a few annotated ground truth sample images are shown below in Figure 2.

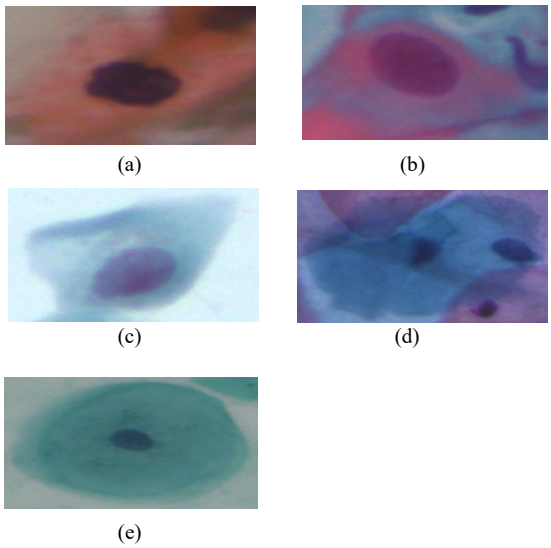


Figure 1: Various Images Of Cervical Cancer From The Initial SIPAKMED Dataset. (A) Dyskeratotic, (B) Parabasal, (C) Koilocytotic, (D) Superficial-Intermediate And (E) Metaplastic

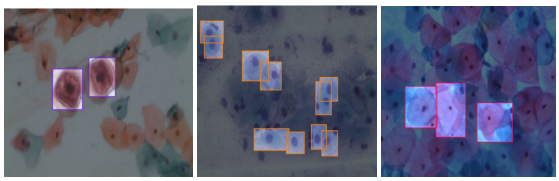


Figure 2: Different Annotated Cervical Cancer Images Of SIPAKMED Dataset. (A) Dyskeratotic, (B) Parabasal And (C) Koilocytotic

3.2 Dataset Augmentation and Image Processing

In order to address the imbalanced class issue, the data augmentation strategy is essential. It helps in balancing the distribution of data. As we can observe from the above Table 2 that, the image numbers belonging to different classes are varying in nature. Data augmentation is used to regenerate the images from the original ones with the help of different parameters like cropping, flipping, rotating. Data augmentation is helpful to enhance the size and quality of training datasets [19]. Each input training image sample is rotated clockwise, counter-clockwise and upside-down giving 3 different output images.

We have considered the following parameters for image processing.

- Resizing an Image

Given that most deep learning architectures need the same size, the original input image has been scaled to 640×640.

- Auto orient of an Image

YOLO models are trained just as the majority of deep learning models, on datasets with consistently right-side up images (for example, all the images in the COCO dataset are upright). Auto-orient is an image preprocessing step that is applied in the dataset and fed to pretrained YOLO models. This ensures that each and every image is physically turned to its proper upright position before being fed to the YOLO model.

3.3 Building a Detection Model using YOLO Network

YOLO is renowned for its quickness and real-time functionality. Because it views object detection as a single regression problem, it can simultaneously predict bounding boxes and class probabilities. Although there have been many variations of YOLO, the key components of this network are mentioned below:

- Backbone Network - Feature Extractor

In order to extract rich, hierarchical information from the input image, this is the main component of the network. Fundamentally, it is a powerful convolutional neural network.

- Neck

The backbone is connected to the forward network further using the neck. To create feature pyramids that enable the model to identify objects

at different sizes (small, medium and large), it combines and enhances the features that the backbone has extracted. For reliable object detection, this combination of multi-scale features is essential. Spatial Pyramid Pooling layer or the Path Aggregation Network is used to increase the information flow between network layers.

- Detection Head

The ultimate predictions are given at the head. For each detected object, it provides the bounding box coordinates, objectness scores (the degree of confidence that an object is present) and class probabilities using the neck's enhanced feature maps.

3.4 Transfer Learning

Due to lack of readily available clinical datasets and necessary computing power, YOLO models are pre-trained using a large-scale, publicly accessible image dataset named COCO dataset intended for object detection and classification and then are fine-tuned with the relevant dataset in specific application domains [20]. The detection head is programmed for COCO dataset's 80 categories. Now using transfer learning concept, we keep the feature extractor and neck elements of the YOLO network unchanged. Only the detection head is modified to detect the classes of the custom dataset. In this way, the COCO dataset's weight and parameters improve the efficacy of tiny target object detection and classification, such as cell nuclei, based on the variable shape of cervical cells.

3.5 Used YOLO Models

The YOLO series is always changing and new iterations offer slight yet noticeable gains in efficiency, accuracy and speed. YOLO's primary concept is to use a neural network to identify objects in a single pass, which makes it real-time. YOLOv5 utilizes a Path Aggregation Network for efficient feature extraction [21-22]. YOLOv8 is not only used for object detection, but also for segmentation and pose estimation capabilities. Such model architecture is used to diagnose in various cancer datasets like colon, skin and cervical [23]. Again, YOLOv9 focuses on optimization of information flow by adding a Generalized Efficient Layer Aggregation Network (GELAN) which enhances feature representation. The release year of the YOLOv10 model is 2024. It eliminates non-maximum suppression (NMS) during inference, in order to get completely end-to-end detection. This makes it suitable for edge devices and applications that require low latency. YOLOv11 offers faster processing rates while maintaining the optimal accuracy to performance ratio due to updated architecture designs and additional enhanced training pipelines. These models are simple to apply in various settings, such as cloud platforms, edge devices and NVIDIA GPU-based systems.

3.6 Software and Hardware Used

The experiments have been implemented with Pytorch (2.2.2) and Torchvision (0.17.2) on a Google Colab Notebook. An i5 11th generation processor and 16GB RAM are part of the hardware system.

4. RESULTS AND DISCUSSION

The analysis and findings related to our investigation is discussed in this section. The performance metrics are mostly based on a comparison of forecasted and real values.

4.1 Object detection models' performance metrics comparison applied on SIPAKMED dataset

A total of seven YOLO neural models are examined in total for the purpose of object detection and classifying cervical cancer cell images. The AdamW optimizer, a learning rate of 0.01 and a momentum value of 0.937 for 100 epochs make up the experimental configuration. The results obtained are given in below Table 4 where input images are applied directly to the model after applying image augmentation and image preprocessing techniques.

Table 4. Comparison of YOLO variant models on SIPAKMED dataset

Name of the model	Precision	Recall	mAP50	mAP50−95	Training Time (in hours)
YOLOv5n	0.800	0.810	0.849	0.653	0.931
YOLOv5s	0.846	0.799	0.852	0.691	1.617
YOLOv5m	0.853	0.802	0.864	0.709	2.798
YOLOv8n	0.800	0.793	0.844	0.656	1.734
YOLOv8s	0.869	0.788	0.866	0.712	1.950
YOLOv8m	0.813	0.817	0.853	0.706	3.259
YOLOv9s	0.848	0.802	0.857	0.702	2.275
YOLOv9m	0.867	0.796	0.861	0.728	3.978
YOLOv10n	0.808	0.807	0.850	0.663	1.272
YOLOv10s	0.832	0.787	0.863	0.692	2.074
YOLOv10m	0.799	0.757	0.820	0.660	3.631
YOLOv11n	0.800	0.818	0.852	0.668	1.001
YOLOv11s	0.855	0.812	0.877	0.713	1.657
YOLOv11m	0.814	0.819	0.836	0.681	3.425
YOLOv12n	0.827	0.825	0.859	0.675	1.490
YOLOv12s	0.856	0.818	0.866	0.711	2.632
YOLOv12m	0.844	0.819	0.871	0.709	4.962
YOLOv13n	0.842	0.829	0.866	0.716	2.588
YOLOv13s	0.856	0.822	0.875	0.739	3.988
Faster YOLOv13s (Ours)	0.859	0.816	0.870	0.710	1.251

A comparative analysis of different fine-tuned transfer learning YOLO models namely YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, YOLOv12 and YOLOv13 are considered. Along with these models, performance of our proposed model named Faster YOLOv13s is also displayed. Based on mAP@50 parameter value, only the best performing variant of all the above YOLO architectures are considered and is presented in Figure 3.

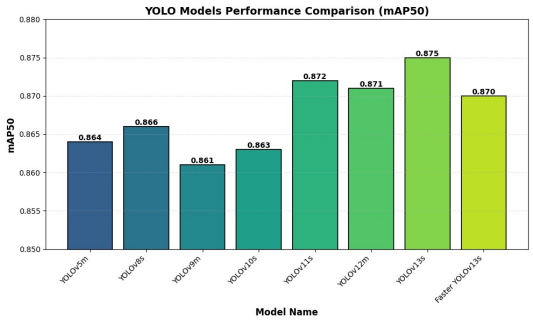


Figure 3: Performance metrics (mAP50) of various object detector models using SIPAKMED dataset

4.2 Performance Metrics Used For Evaluation Of The Various Object Detection Models

The ratio of the intersection of both boxes and their union is used to compute the intersection over union (IoU) of the predicted bounding boxes with regard to the ground truth boxes. Most majority of evaluation criteria for detection methods are predicated on this [24]. Following Equation 1 can be utilized to calculate the IoU between two arbitrary points, A and B. A higher value of IoU means a better prediction.

$$IoU = A \cap B / A \cup B \quad (1)$$

where, the intersection or the region where the ground truth box and the forecasted box overlap is the numerator. It displays the common shared area between both boxes.

The area covered by both boxes together is the denominator component. It is the sum of the areas of the ground-truth box and the anticipated box and then subtracting the intersection area to avoid counting the overlap twice. By setting a preset IoU threshold, the algorithm's true positive (TP), false positive (FP) and false negative (FN) detections can

be determined. This makes it possible to estimate well-known machine learning parameters like recall and precision. Below, Equations 2 and 3 represent the precision and recall formulas.

$$\text{Precision} = TP / TP + FP \quad (2)$$

where, false positives are represented by FP and true positives by TP.

$$\text{Recall} = TP / TP + FN \quad (3)$$

where, true positives are denoted by TP and false positives by FN.

The precision values for each recall level discovered during the evaluation process are averaged to produce a number of AP values, which are then used to calculate mean average precision (mAP). To get this value, the mean of the obtained AP values for all object categories is calculated. Each AP is linked to a unique IoU threshold in the 0.5 to 0.95 range by using a step of 0.5. We indicate the mean average precision by the following Equation 4.

$$mAP = 1/n \sum_{k=1}^{n} AP_k \quad (4)$$

where, APk is the AP of class k and n is the number of classes.

4.3 Key components in proposed model

The suggested model has the backbone frozen, while keeping only the neck and head trainable. The two primary components included in this model are- C2f module with integrated deformable attention that enhances feature extraction in the neck. The other one is Deformable attention block along with FFN with residual connections is a complete transformer like block used for attention. Below Table 5 displays the comprehensive components comparison mainly in the "neck" of the models.

Table 5. Architectural components comparison of YOLOv13 variants

Component (Neck)	Pretrained YOLOv13n	Pretrained YOLOv13s	Faster YOLOv13s
Core module	C2f standard	C2f standard	C2fDeformable
Feature Aggregation	FullPAD tunnels	FullPAD tunnels	FullPAD+Deformable Attention
Multi-scale Fusion	HyperACE mechanism	HyperACE mechanism	HyperACE+Deformable sampling
Attention mechanism	Hypergraph-based	Hypergraph-based	Hypergraph+Deformable Attention
Spatial Pooling	SPPF (Spatial Pyramid Pooling Fast)	SPPF	SPPF

4.4 Performance Metrics Of Best Performing Model

Out of all the models that have been investigated, it is discovered that Faster YOLOv13s model is quite competitive in all respects of the works mentioned in Table 1 and reported by all the authors. It has achieved mAP_{50}, mAP_{50-95} values as high as 87.00% and 71.00% respectively and taking an overall training time of only 1.251 hours. We have examined three variants named YOLOv13n, YOLOv13s and Faster YOLOv13s based on their performance in important measures such as mean Average Precision (mAP), training time and model parameters. Faster YOLOv13s has achieved an outstanding training efficiency, taking only 1.251 hours versus 3.988 hours for standalone YOLOv13s, a 68.60% reduction in training time. Faster YOLOv13s outperforms the lightweight YOLOv13n by 51.70% in training time, making it the most efficient variant. Faster YOLOv13s has 6.08M trainable parameters, which is in between the ultra-lightweight YOLOv13n (2.49M) and YOLOv13s (9.53M). Below Figure 4 depicts the same. This balanced parameter count provides sufficient model capacity while avoiding the computational burden. Though Faster YOLOv13s has somewhat lower mAP_{50} value than YOLOv13s, the performance degradation is negligible. Given that accuracy should not be the exclusive criterion for evaluating a deep learning model, Faster YOLOv13s model's results are quite promising due to its fast inference time and lightweight design.

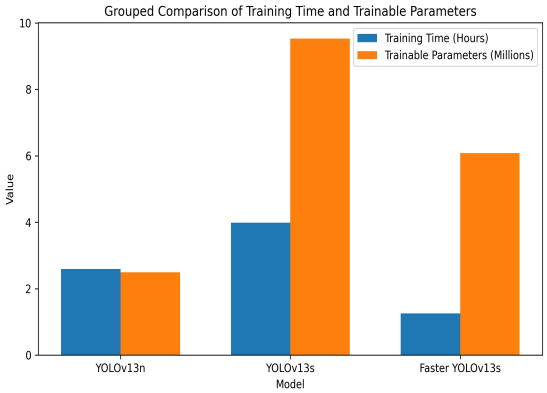


Figure 4: Grouped Comparison of Training Time and Trainable Parameters

The confusion matrix generated from the Faster YOLOv13s model is given below Figure 5. It shows the true labels and predicted labels are identical. Hence, we conclude that this model has almost all accurate predictions.

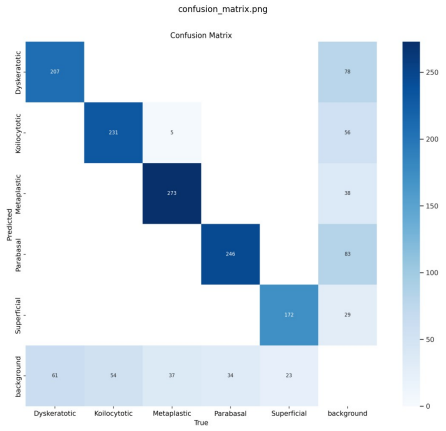


Figure 5: Confusion matrix of proposed model using SIPAKMED dataset

The precision and recall graphs for all the classes of the dataset using Faster YOLOv13s model are given in Figure 6. Plotting precision versus recall for various thresholds is shown on the graph. The broad blue line shows the PR curve averaged over all classes, whereas each coloured line shows the PR curve for a particular class. The mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 is known as mAP@0.5. It is a widely used metric for identifying objects. mAP_{50} value achieved using Faster YOLOv13s is 87.00%. This model performs well across all classes, as seen by the high total score and gives high accuracy in object localization and classification tasks.

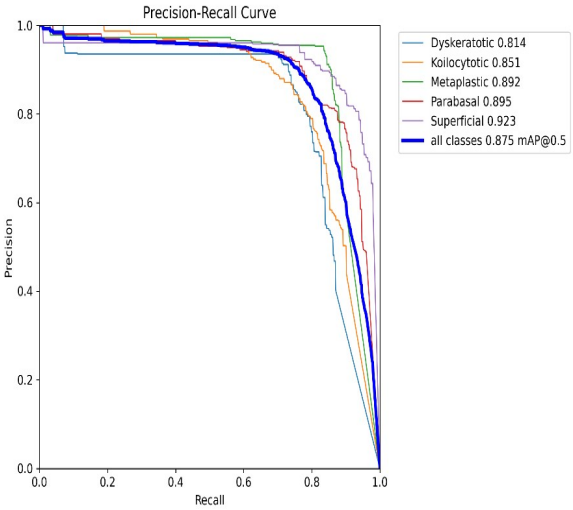


Figure 6: Precision and Recall graphs across all classes

4.4 Analysis Of Predicted Images Using Best Performing Model

Figure 7 displays the actual labels or ground truth images. Whereas, Figure 8 displays the corresponding forecast and confidence score along with the result image. A bounding box encircling the cell nucleus is predicted by the YOLO model. The outcomes demonstrate that our approach is able to distinguish between cells and other impurities. In addition, it is able to detect multiple classes present in the image of the overlapped cells.

The data.yaml file used in our experimental setup contains all the class names corresponding to the class-IDs in annotation files. Below Table 6 gives the details of the yaml file. When a single class image is given input, then our model has predicted the class name along with the bounding box with high confidence score.

Table 6: Details of yaml file

Class Name	Class ID
Dyskeratotic	0
Koilocytotic	1
Metaplastic	2
Parabasal	3
Superficial-Intermediate	4

Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

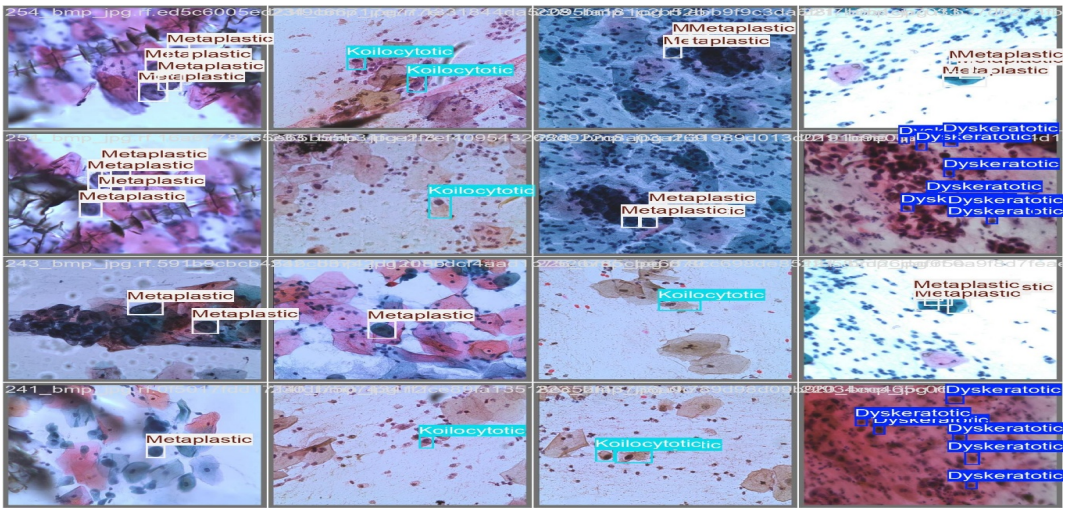


Figure 7: Ground Truth Labels Of Sample Images

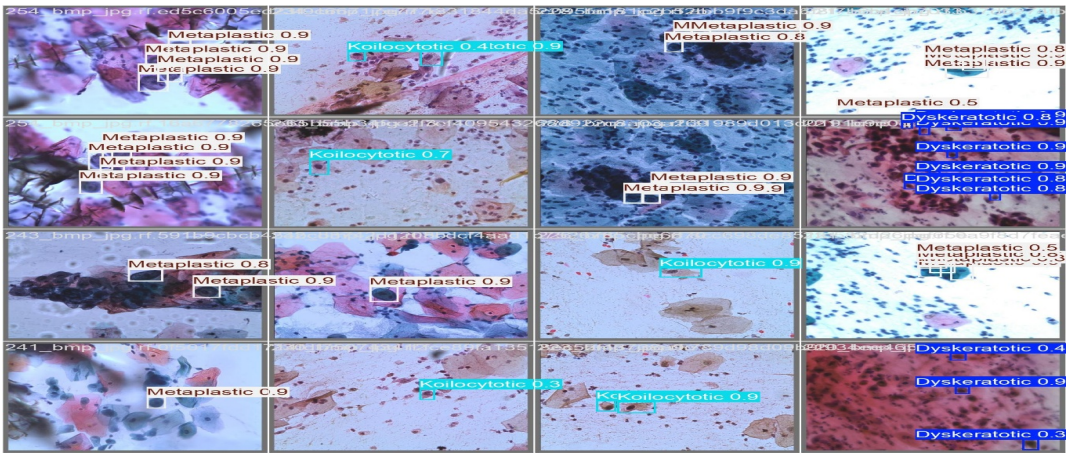
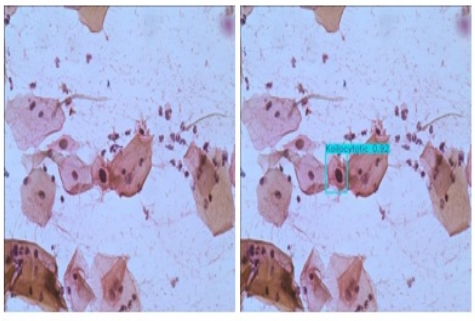


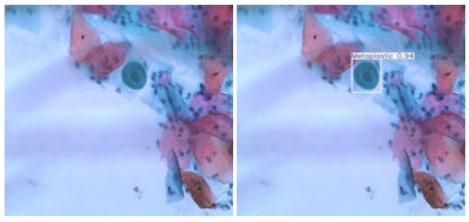
Figure 8: Predicted Class And Confidence Score Using Proposed Model

The following Figure 9 and Figure 10 shows that the predicted image label and the actual image label match with high accuracy.



(a) Actual(Class ID:1) (b) Predicted

Figure 9: Single Image Prediction For Koilocytotic Class



(a) Actual(Class ID:2) (b) Predicted

Figure 10: Single Image Prediction For Metaplastic Class

4.5 Comparison of best performing model with other related works

As per below Table 7 details, our suggested Faster YOLOv13s model shows a definite and significant performance benefit based on a

comparative examination of current object-detection techniques assessed using the SIPAKMED dataset. While more recent architectures like YOLOv5l (Lohith *et al*., 2023) and RF-DETR (Sapkota *et al*., 2025) reported improved mAP50 values of 59.00% and 69.00%, earlier techniques, such as Faster R-CNN variants reported by Ren *et al*., 2015 and Sampaio *et al*., 2021 achieved mAP_{50} scores of 64.50% and 33.70% respectively. Our proposed Faster YOLOv13s framework, on the other hand, achieves a mAP_{50} of 87.00%, which is competent enough when compared with other models. The key initiative taken in our approach is achieving competent results even after significant reduction in number of trainable parameters and considerably taking lesser amount of training time. Its remarkable training efficiency, balanced parameter count and competitive detection accuracy make it the ideal model for deployment scenarios in which both performance and computational efficiency are crucial. This higher detection accuracy demonstrates the robustness and improved discriminative ability of our model for cervical cell detection on the SIPAKMED dataset, showing that it is more effective than current-state-of-the-art methods at capturing fine-grained morphological features and managing dataset variability.

Table 7: Comparison with related works

Author Name	Model Name	mAP_{50}
Ren *et al.* (2015) [6]	Faster-RCNN	64.5
Sampaio *et al.* (2021) [12]	Faster RCNN	33.70
Lohith *et al.* (2023) [13]	YOLOv5l	59.00
Sapkota *et al.* (2025) [9]	RF-DETR	69.00
Our work (2025)	Faster YOLOv13s	87.00

5. LIMITATIONS

Since the work is on medical domain, accurate diagnosis plays a vital role in detecting the cancer cells. From the experimental results, it is found that the mAP_{50} value is improved by a huge percentage than the SOTA architecture results. Still, other cervical cancer datasets can be examined to determine the robustness of the suggested model.

6. CONCLUSION AND FUTURE WORK

In this paper, we use the object detection method to construct a cervical cell diagnosis system. Knowing the exact location of different cervical cells is crucial for pathologists. Bounding boxes are drawn around these cells using object detection, which helps the pathologist focus on abnormal spots on a slide. Diagnostic time and effort can be greatly reduced by using object detection to highlight particular cells or regions rather than manually screening an entire slide with a few thousands of cells. Frequently, object detection and classification go hand in hand to enhance efficiency, accuracy and objectivity of cervical cancer diagnosis.

In the future, cell characteristics such as nucleus size, shape, nuclear-to-cytoplasmic ratio, and overall cell size can be recovered and subjected to a more thorough analysis to check the severity of cervical cancer. However, in contrast to every object detection model covered in literature review, in terms of both efficiency and computational resources, our suggested Faster YOLOv13s model performed better than all other models. Still more work can be done to decrease the memory size and increase the inference time with the help of quantization and LoRA techniques. Hence a fully automated cervical cell-based detection system can be made by modifying the model to work on limited resources and edge devices.

ACKNOWLEDGEMENTS

The first author extends deep gratitude to Dr. Rajdeep Chatterjee and Dr. Mahendra Kumar Gourisaria for their constant and unwavering support, without which this study would not have been possible.

REFERENCE:

[1] Who Facts. Available: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer. Accessed on: 15/05/2025
[2] Dominck cunningham yaffa. Available: https://www.pbglaw.com/blog/how-accurate-is-the-pap-test. Accessed on: 17/04/2025
[3] Turashvili G., HPV associated cervical squamous cell carcinoma. Available: https//www.pathologyoutlines.com/topic/cervixSCC.html. Accessed on: 17/04/2025
[4] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object

detection. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 779-788).
[5] Wang Y, Yang C, Yang Q, Zhong R, Wang K, Shen H. Diagnosis of cervical lymphoma using a YOLO-v7-based model with transfer learning. Scientific Reports. 2024 May 14;14(1):11073.
[6] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems. 2015;28.
[7] Bochkovskiy A, Wang CY, Liao HY. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020 Apr 23.
[8] Li X, Li Q. Detection and classification of cervical exfoliated cells based on faster R-CNN. In 2019 IEEE 11th international conference on advanced infocomm technology (ICAIT) 2019 Oct 18 (pp. 52-57). IEEE.
[9] Sapkota R, Cheppally RH, Sharda A, Karkee M. RF-DETR Object Detection vs YOLOv12: A Study of Transformer-based and CNN-based Architectures for Single-Class and Multi-Class Greenfruit Detection in Complex Orchard Environments Under Label Ambiguity. arXiv preprint arXiv:2504.13099. 2025 Apr 17.
[10] Li X, Li Q. Detection and classification of cervical exfoliated cells based on faster R-CNN. In2019 IEEE 11th international conference on advanced infocomm technology (ICAIT) 2019 Oct 18 (pp. 52-57). IEEE.
[11] Xiang Y, Sun W, Pan C, Yan M, Yin Z, Liang Y. A novel automation-assisted cervical cancer reading method based on convolutional neural network. Biocybernetics and Biomedical Engineering. 2020 Apr 1;40(2):611-23.
[12] Sampaio AF, Rosado L, Vasconcelos MJ. Towards the mobile detection of cervical lesions: a region-based approach for the analysis of microscopic images. IEEE Access. 2021 Nov 8;9:152188-205.
[13] Lohith M, Bardhan S, Bandyopadhyay O. Cervical Pap Smear Screening and Cancer Detection Using Deep Neural Network. In Current Applications of Deep Learning in Cancer Diagnostics 2023 Feb 22 (pp. 125-137). CRC Press.
[14] Kalbhor M, Shinde S, Wajire P, Jude H. CerviCell-detector: An object detection approach for identifying the cancerous cells in pap smear images of cervical cancer. Heliyon. 2023 Nov 1;9(11).
[15] Ontor MZ, Ali MM, Ahmed K, Bui FM, Al-Zahrani FA, Mahmud SH, Azam S. Early-stage cervical cancerous cell detection from cervix images using yolov5. Computers, Materials and Continua. 2023;74(2):3727-41.
[16] Wu XJ, Zhao CJ, Meng C, Wang H. Research on Cervical Cancer p16/Ki-67 Immunohistochemical Dual-Staining Image Recognition Algorithm Based on YOLO. arXiv preprint arXiv:2412.01372. 2024 Dec 2.
[17] Indugu VV, Sai D. Advanced cervical cancer detection using YOLOv5. Security Issues in Communication Devices, Networks and Computing Models: Volume 2. 2025 May 8:197.
[18] Sipakmed dataset. Available: https://www.kaggle.com/datasets/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed. Accessed on: 05/11/2025
[19] Abdulkareem IM, AL-Shammri FK, Khalid NA, Omran NA. A Proposed Approach for Object Detection and Recognition by Deep Learning Models Using Data Augmentation. International Journal of Online & Biomedical Engineering. 2024 May 1;20(5).
[20] Mammeri S, Amroune M, Haouam MY, Bendib I, Corrêa Silva A. Early detection and diagnosis of lung cancer using YOLO v7, and transfer learning. Multimedia Tools and Applications. 2024 Mar;83(10):30965-80.
[21] Lv D, Yi L, Liu L, Chen Y, Chen X, Liu R. YOLO-TCT: An Effective Network For Long-Tailed Cervical Cell Detection. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2025 Apr 6 (pp. 1-5). IEEE.
[22] Peng Z, Hu R, Wang F, Fan H, Eng YW, Li Z, Zhou L. Deep adaptively feature extracting network for cervical squamous lesion cell detection. InInternational Conference on Machine Learning for Cyber Security 2022 Dec 2 (pp. 238-253). Cham: Springer Nature Switzerland.
[23] Palanivel N, Deivanai S, Sindhuja B. The art of YOLOv8 algorithm in cancer diagnosis using medical imaging. In 2023 International Conference on System, Computation, Automation and Networking (ICSCAN) 2023 Nov 17 (pp. 1-6). IEEE.
[24] Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 658-666).