Heliyon 9 (2023) e21388

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Mask region-based CNNs for cervical cancer progression diagnosis on pap smear examinations

Carolina Rutili de Lima [a,*], Said G. Khan [b], Syed H. Shah [c], Luthiari Ferri [d]

[a] Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan
[b] Department of Mechanical Engineering, College of Engineering, University of Bahrain Isa Town, Bahrain
[c] College of Electrical and Communication Engineering, Yuan Ze University, Taoyuan, Taiwan
[d] Instituto de Patologia de Passo Fundo, Ijuí, Brazil

ARTICLE INFO

Keywords:
Cervical cancer
Mask RCNN
Deep learning
Cells segmentation and classification
Whole tissue classification
Health and technology

ABSTRACT

This research presents a novel approach for cervical cancer detection and segmentation using tissue images with multiple cells. The study employs a novel deep learning architecture based on Mask Region-Based Convolutional Neural Network (RCNN) and statistical analysis. This new architecture enables us to achieve a high percentage of detection and pix-to-pix area segmentation. A mean Average Precision (mAP) higher than 60% for 3-class and 5-class was achieved. In addition, higher F1-scores of 70% for 3-class and 5-class were obtained. This investigation is a collaborative work, where a medical consultant collected the samples from the Papanicolaou (Pap) Smear examination and labeled the cells presented to the liquid-based cytology (LBC). Furthermore, the online available benchmark data set, *SIPaKMeD*, was also utilized. Additionally, sample images from the *Mendeley* data set were also labeled by the trained medical consultant for comparison. The proposed scheme automatically generates a full report for a medical consultant to identify the location of the malicious cells in the given images and expedite the diagnosis and treatment process.

1. Introduction

Cervical cancer remains one of the most prevalent malignancies in women, particularly in emerging countries with limited access to institutional health systems. [1]. Even though vaccination is widespread nowadays, most women remain vulnerable because vaccines are not 100 percent effective against all the variants of the viruses. Considering this perspective, the problem of assisting all individuals affected by cervical cancer becomes incredibly challenging for a public health system, particularly given the number of women who are already affected and those who may be carrying the virus but are unaware of it.

It is also a fact that most low-income countries and emerging countries don't have the capacity to vaccinate all women against cancer. In addition, they do not have a proper preventive mechanism and lack awareness to prevent the transmission of the cervical cancer virus. Cancer treatment options are also very limited in these countries. Globally, any public health system will have difficulty coping with many women who are already suffering from cervical cancer and many virus carriers. However, as noted by the World Health Organization (WHO), most of the merging countries that cannot vaccinate, educate, and prevent the transmission of the virus in their populace are more vulnerable [1].

* Corresponding author.
E-mail address: 80975004h@gapps.ntnu.edu.tw (C. Rutili de Lima).

https://doi.org/10.1016/j.heliyon.2023.e21388
Received 20 June 2023; Received in revised form 19 October 2023; Accepted 20 October 2023

Available online 25 October 2023

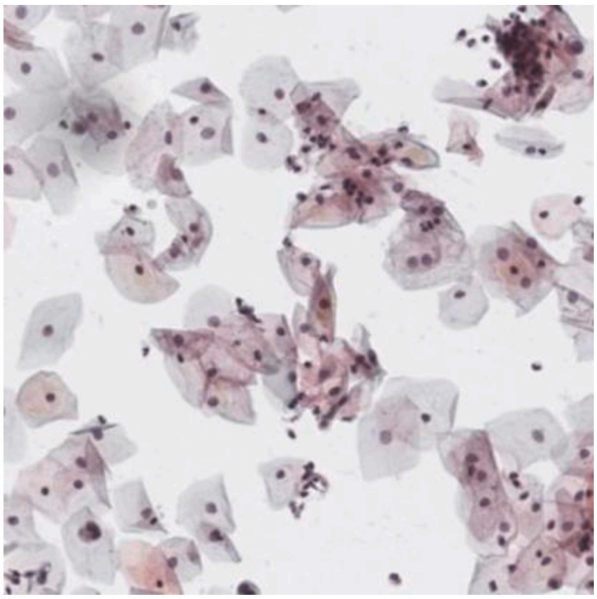


Fig. 1. Tissue cell example.

It is also worth mentioning that globally, there is a shortage of consultant pathologists in the field of analyzing cells and interpreting pap smear tests [1]. Therefore, this is one of the major reasons for the slow rate of cervical cancer diagnosis in women [2]. Usually, the examination of a patient requires at least two specialists, i.e., a gynecologist and a consultant pathologist. However, available experts in the field are usually extremely occupied as they have to deal with multiple assignments.

The pap smear test is the most common way to detect cervical cancer. During this examination, the gynecologist collects a sample of the cells in the cervix [3,4]. After the sample collection, the sample is sent to a pathologist for analysis. However, this analysis can take hours because of the huge number of cells in one human tissue. Also, it may be necessary to examine more than one tissue, each one containing dozens of cells, being very exhaustive for the doctor consultant. In addition, an expert may not be available on-site, and the samples may need to be transported to another far-off destination for analysis, which further delays the diagnosis process [1].

For precautions, some specialists recommend that women should do the test once a year for prevention and early treatment [5]. Nonetheless, it remains challenging to investigate the tissue by specialists for various reasons. The Brazilian pathologist who helped us during this research mentioned the need for a software-based tool that could show the location of the cell that may be affected or have some degree of cancer, which will significantly improve the speed of analysis and detection process. Furthermore, if there's software with high reliability, we could utilize it for early treatment and diagnosis of women with the disease.

Nevertheless, artificial intelligence (AI) techniques such as deep learning are becoming increasingly popular in solving problems in diverse areas, e.g., green power, face detection for security applications, social media, and image processing in medical diagnosis. One of the hot areas in image processing is cancer diagnosis, and AI has been used to explore the great potential of automatic cancer diagnosis and many other diseases. In the near future, AI-based sophisticated software will be available to aid consultants in this field as referred in [6]. One of the challenges in this field is when the cells overlap each other, as shown in Fig. 1. This makes it even harder for Deep Learning algorithms to distinguish, classify, and detect different types of carcinogenic cells, yet different laboratories use different colors, which makes it difficult to employ the same diagnostic and detection tools.

Regardless, to make that happen, a large database of accurately labeled images will be required to train a deep-learning NN; after training the deep NN, it should be tested and validated with the unseen images. The basic idea of training NN and testing is shown in Fig. 2. Once satisfied with the performance, it would be deployed in practice. However, there is still no universal and highly robust solution for cancer diagnosis, and without an expert medical consultant, the results may lead to false positives or false negatives. A significant amount of research has been conducted in this area to train deep learning schemes efficiently to classify an area of an image. New improvements are reported worldwide, eventually leading to automatic cancer detection and diagnosis software tools.

Also, whole tissue cells are cells that are studied within their natural tissue environment, surrounded by neighboring cells, extracellular matrix, and other cell types that make up the tissue microenvironment. The micro-environment can impact the behavior and functionality of these cells [7] [8]. In contrast, a single cell refers to a cell that has been removed from its original tissue environment and cultured in vitro, usually in a laboratory [7]. For our research, we opted to analyze a whole tissue slide containing multiple cells rather than a single cell, as many studies in this area have already been conducted.

Examining the current literature review, cited in the next chapter (Background and Literature Review), we observed that the current work in cervical cancer using whole tissue cells hasn't yet been fully automatized for detection, classification, and segmentation. The single crop cell classification and segmentation, or the weak mean Average Precision (mAP), are not enough to deliver a tool to be used by pathologists. To further improve the investigations in this area, we propose a model in this research to automatize the

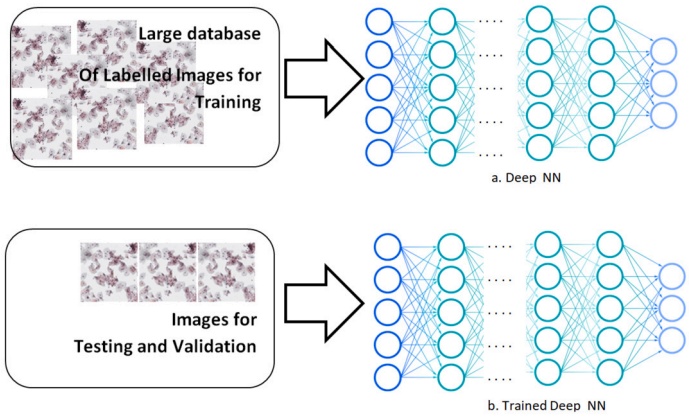


Fig. 2. Deep NN-based Image Analysis.

images collected from the microscopy and generate a report to assist the consultant pathologist. So we presented the following main contributions (and novelties) of this paper:

- A new method for a whole tissue slide cell classification, segmentation, and detection using a modified Mask RCNN ResNeXt backbone, proven by using statistical variables, such as mAP (median average precision), mAR (median average recall), and F1-score;
- Currently, most of the state-of-the-art cervical cancer diagnostic/detection schemes are not fully automatic to deliver a full report. In our case, a comprehensive report is automatically generated to be seen by a medical consultant, which expedites the diagnosis of the malignant cells;

The remaining paper has been divided into six sections. In Section 2, the Background and Literature Review are presented. In Section 3, the Image Data sets and the Methodology are discussed. In Section 4, a detail of the experiments carried out. In Section 5, the discussions of our results are provided. We finalize the paper with the conclusion and possible next steps for our research in Section 6.

2. Background and literature review

In the past ten years, many researchers have investigated this field. In this section, we highlight some of the most important ones and their main contributions to cervical cancer detection using Artificial Intelligence (AI). In the research work of Ghoneim et al. [9] and Waly et al. [10], deep learning networks were proposed for feature extraction (i.e., cell detection), and the Extreme Learning Machine (ELM) is used as the classifier for each cell detected. Yet, in [9], a CNN (Shallow, VGG-16-Net, and CaffeNet) was implemented and then connected to two ELMs (removing the softmax layer) for classification. The output of the first ELM is set to give normal or abnormal cells, and the other one is set to give classes of normal and abnormal cases.

Furthermore, Waly et al. [10] reached the highest model accuracy of 97.96% (7-class). The author employed the IDCNN-CDC model, which includes four major processes: 1) Preprocessing, the Gaussian Filter (GF) is applied to enhance data by removing noise from the dataset; 2) Segmentation, Tsallis entropy with the dragonfly optimization (TE-DFO) to make it easier to identify the malicious portions; 3) Feature extraction, the dataset images are fed into SqueezeNet; 4) Classification, the extracted features are used to the weighted extreme learning machine.

Win et al. [11], Rehman et al. [12] and Sabeena and Gopakumar [13] applied deep learning algorithms followed by machine learning techniques to classify cervical cancer cells. On the other hand, [11] employed a bagging ensemble classifier that computes the output of base learners during the classification stage. Moreover, Rehman et al. [12] and Sabeena and Gopakumar [13] tested each model separately at the end.

Similar to the work by Jia et al. [14], Yaman and T. Tuncer [15] also used SVM as the last step in the architecture. However, the CNNs applied were different. The first step was to use DarkNet19 and DarkNet53 in a “Pyramid Deep Feature Extraction Model”, which extracts features in three distinct sizes: 128x18, 64x64, and 32x32. Following that, the features have been merged, and NCA is used to determine the 1000 most relevant features. In the last step, SVM is used to classify them into 4 classes. Manna et al. [16] and Pramanik et al. [17] implemented two different methodologies using a Fuzzy Learning ensemble from three different CNNs. Both studies pre-trained on the ImageNet dataset. Moreover, despite using different strategies, they got good accuracy. Hussain et al. [18] also used the CNNs and ensembled with Resnet-50, Resnet-101, and GoogleNet. The ensemble classifier seeks the maximum number of classifiers’ decisions and weighs them simultaneously to improve efficiency and performance, and it selects a class based on the highest number of votes received.

Chen et al. [19,20] and Tan et al. [21] focused on the whole slide image (WSI) through classification using different architectures. [19] has three main parts: 1) segmentation, which extracts cells from the whole slide image (WSI); 2) classification, a new visual

Table 1
Private Dataset.

Cell Name	Cancer/Normal	Cells Quantity
SSE (Superficial Squamous Epithelial)	Normal	431
ISE (Intermediate Squamous Epithelial)		976
CE (Columnar Epithelial)		3
Endocervical		10
Mild SNKD (Mild Squamous non-keratinizing dysplasia)	Pre cancer - Abnormal	297
Moderate SNKD (Moderate Squamous non-keratinizing dysplasia)		265
Severe SNKD (Severe Squamous non-keratinizing dysplasia)	Cancer - Abnormal	163
SCCIS (Severe cell carcinoma in situ intermediate)		

geometry group (VGG) called CompactVGG that is faster than the original one; 3) human aided visualization, providing two visual display modes for users to review and modify. In the work by [20], the architecture has three models: 1) LR model, it's fed with cropped images of 512 × 512 pixels from the WSI; HR model has an input image of 256 × 256 cropped according to the location heatmap and outputs a new lesion probability; RNN integrates the top 10 lesion cells and outputs their probabilities. In another instance, i.e., the work by Tan et al. [21], the Faster RCNN extracts the image information to get the feature map and the region proposal network, so at the end, it's possible to have access to the target category together with the target location. The WSI was zoomed in 200X and cropped to feed the CNN, and different images were also enhanced and augmented.

In contrast to previous research work, [22] apply Mask RCNN for the dataset images that contain just cells cropped and get their segmented image. For training and testing, these images are fed to another algorithm (Visual Geometry Group-like Network). In the work by [23], the authors investigated the possibility of a mobile-based framework detecting cervical lesions. This framework is based on the Internet of Things (IoT) that integrates the μSmartScope for the acquisition of the sample images with the deep learning model for detection and classification. They used the SIPaKed Dataset for training and then testing on their framework. The best performance was Faster-RCNN using the five classes in SKIPaKed. Moreover, the authors tried to push these models in the data acquired by the μSmartScope. However, the outcome was limited.

Xiang et al. [24] is also a research based on CNN detection and classification of cervical cancer tissue cells in different datasets. The architecture starts with the Darknet-53 network trained on Imagenet, used as a feature extractor, and then fine-tuning all convolution layers of YOLO3. The authors also compared different networks (FasterR-CNN, YOLOv3 416, YOLOv3 608, Tiny YOLOv3), and the one that achieved the best was YOLOv3 608, mAP of 0.574.

Many researchers in this field have focused on reaching a higher classification average using cropped images and with two steps: 1) feature extraction and 2) classification. Moreover, the ones that are targeting the segmentation process don't deliver a report for consulting the cells and helping the medical consultant in this area.

3. The image data and methodology

This section presents the essential resources, such as image data sets employed for this work. In addition, the statistical metrics and the architecture/topology are discussed in this section.

3.1. The dataset

This study utilized three different image data sets, which are described in detail in the following subsections:

3.1.1. Private dataset

A doctor partnered with a Brazilian clinic situated in Ijuí/Rio Grande do Sul (Instituto de Oncologia de Ijuí) to collect during the past years a private dataset of Pap Smear samples with various colorations taken from different microscopy, including some stored images from previous examinations. The doctor responsible for the clinic reviewed and examined all the data, ensuring that sensitive information such as names and ages were not included.

The coauthor (Pathologist) obtained images of tissue cells using microscopy and labeled them according to the Bethesda system. The classification of these cells can be found in Table 1, and examples of each cell's classification are provided in Table 2.

For those images, we cropped them into smaller ones to reduce the size (high resolution and also the number of cells in each make these images very large). Moreover, the annotations were also cropped according to the images, and then all of these were uploaded to roboflow.com for pre-processing. The Roboflow website is a very handy tool for pre-processing and augmentation of a large number of images.

3.1.2. Open source dataset SIPaKMeD

This dataset is a benchmark, and it can be downloaded online: link, it contains a total of 4049 images, classified into 5 different cells' categories, and published data was in 2018 [25]. The number of cells of each class is given in Table 3, and an example of a cell's classification is presented in Table 4.

Table 2
Private dataset example cells.

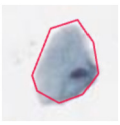
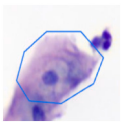
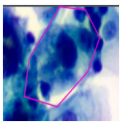
SSE (Normal)	ISE (Normal)	CE (Normal)	Endocervical (Normal)	Mild SNKD (Pre Cancer)	Moderate SNKD (Pre Cancer)	Severe SNKD (Cancer)	SCCIS (Cancer)
							

Table 3
Open Source Dataset SIPaKMeD.

Cell Name	Cancer/Normal	Cells' Quantity
Superficial-Intermediate	Normal	831
Parabasal		787
Metaplastic	Pre cancer - Abnormal	793
Koilocytotic		825
Dyskeratotic	Cancer	813

Table 4
Open source dataset SIPaKMeD example cells.

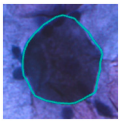
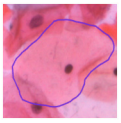
Dyskeratotic (Cancer)	Koilocytotic (Pre-Cancer)	Metaplastic (Pre-Cancer)	Parabasal (Normal)	Superficial-Intermediate (Normal)
				

Table 5
Open source dataset Mendeley.

Cell Name	Cancer/Normal	Cells' Quantity
SSE (Superficial Squamous Epithelial)	Normal	1
ISE (Intermediate Squamous Epithelial)		2
CE (Columnar Epithelial)		0
Endocervical		1
Mild SNKD (Mild Squamous non-keratinizing dysplasia)	Pre cancer - Abnormal	77
Moderate SNKD (Moderate Squamous non-keratinizing dysplasia)		323
Severe SNKD (Severe Squamous non-keratinizing dysplasia)	Cancer - Abnormal	121
SCCIS (Severe cell carcinoma in situ intermediate)		

However, the annotations for this dataset are in a '.dat' and cannot be fed directly into AI algorithms. For this reason, these were converted to '.json' annotations. After conversion, these images were also uploaded to the Roboflow website to organize the image data properly.

3.1.3. Open source dataset Mendeley

The Mendeley dataset ([26]) was also utilized in this study, which was published online in 2018 and is still available for downloads. This repository comprises a total of 963 images. The pap smear images were captured at 40x magnification using a Leica ICC50 HD microscope and were collected and prepared using the liquid-based cytology technique from 460 patients.

However, the Mendeley dataset was found to be deficient in terms of annotations. To rectify this, one of the co-authors (a medical consultant) labeled some images to be used, which are included in the dataset. The labeling was performed according to the first dataset annotations. This dataset was employed because the number of malicious cells was insufficient compared to the other classes. Table 5 presents the cell number used in this dataset.

3.1.4. Merged dataset

In order to obtain optimal training results, all the datasets were combined into a bigger one with more samples, as seen in Table 6. This combination of datasets allowed the network algorithm to process the features of each class more effetely since it will have more data to learn from it.

Table 6
Merged datasets - Cells' quantity and classes.

Classification 1	Cancer/Normal	Classification 2	Cells' Quantity
SSE (Superficial Squamous Epithelial)	Normal	Superficial-Intermediate	2241
ISE (Intermediate Squamous Epithelial)			
CE (Columnar Epithelial)		Parabasal	801
Endocervical			
Mild SNKD (Mild Squamous non-keratinizing dysplasia)	Pre Cancer - Abnormal	Metaplastic	1167
Moderate SNKD (Moderate Squamous non-keratinizing dysplasia)		Koilocytotic	1413
Severe SNKD (Severe Squamous non-keratinizing dysplasia)	Cancer - Abnormal	Dyskeratotic	1096
SCCIS (Severe cell carcinoma in situ intermediate)			

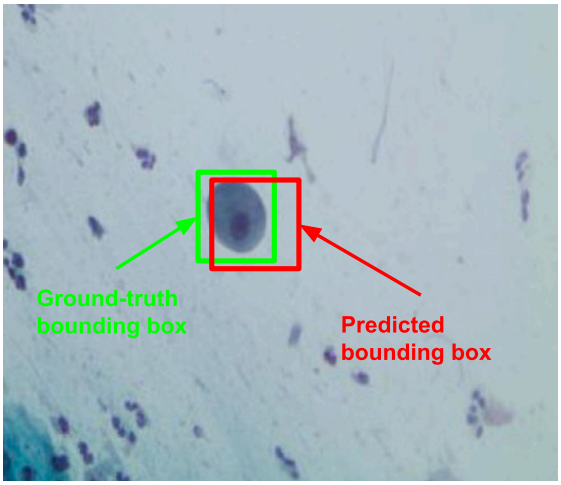


Fig. 3. IoU example.

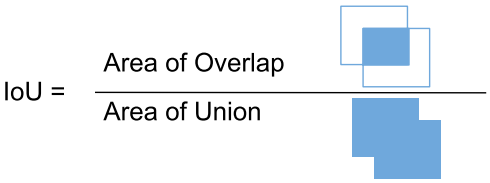


Fig. 4. IoU overlap ratio.

3.2. Statistical metrics

A set of statistical parameters is employed to evaluate the proposed architecture's efficiency compared to prior research, and these parameters will e shown in this section, such as mAP, mAR, and F1- score. In the task of cell segmentation and localization, the objective is to identify the class to which each cell belongs and locate all its pixels. Hence, relying solely on metrics such as Precision and Recall may not be suitable. For the following variable descriptions, the work proposed by Sharma et al. ([27]) is used as a reference in this study for the purpose of clarity.

3.2.1. Intersection over Union (IoU)

The Intersection over Union (IoU) metric is a widely used evaluation measure in computer vision tasks, particularly in object detection and segmentation. It quantifies the extent of overlap between a predicted bounding box and the ground-truth bounding box as demonstrated in Fig. 3 ([27]).

The computation of IoU involves calculating the area of overlap between the two bounding boxes and then dividing it by the area of their union. Fig. 4 demonstrates the process of computing the IoU metric, as described by [27]. The IoU metric is an important tool for evaluating the accuracy of object detection and segmentation models, as it provides a direct measure of the degree of overlap between the predicted and ground-truth bounding boxes.

3.2.2. Precision and recall

The prediction model, for example, for five images, will give us: [0.1, 0.5, 0.7, 0.95, 0.34]. We also need to set a threshold, a hyperparameter that we can change. If we put it as 0.6, it means that if the model predicts above or equal to 0.6, then the prediction belongs to its class; otherwise, it does not.

If we consider that, we are predicting cars. We could get: [not_car, not_car, car, car, not_car].

Before approaching Precision and Recall, we need to consider also these variables:

- True Positive (TP): number of times the model predicted the positive input sample correctly (i.e., car) as Positive;
- False Positive (FP): number of times the model incorrectly predicted the negative sample (i.e., no_car) as Positive;
- True Negative (TN): number of times the model correctly predicted the negative sample (i.e., no_car) as Negative;
- False Negative (FN): number of times the model incorrectly predicted the positive input (i.e., cat) as Negative.

Another important matter, cited in [28], is that precision measures the accuracy of the positive predictions made by a classifier, while recall assesses the ability of the classifier to capture all positive instances in the dataset. We need to know how to balance these trade-offs between these variables.

3.2.3. Precision

Precision in Deep Learning is a metric used to evaluate the accuracy of a model's positive predictions. It measures the proportion of correctly classified samples that were identified as positive by the model out of all the samples that were classified as positive (both correct and incorrect). This metric is particularly useful in Deep Learning applications where the goal is to identify a specific class among multiple classes, such as object detection in computer vision. By computing precision, we can determine how well the model is able to distinguish between positive and negative samples and make accurate predictions.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Equation (1) is the precision representation in terms of true positives and false positives.

Furthermore, [29] underscores the significance of precision, highlighting its critical role in assessing the quality of search outcomes. High precision s ignifies that a retrieval system predominantly delivers pertinent results, whereas low precision implies the inclusion of numerous irrelevant documents in the retrieved set.

3.2.4. Recall

Recall refers to the ability of a model to correctly identify all instances of a certain class, such as positive or negative samples. The recall is calculated by dividing the number of correctly classified positive samples by the total number of positive samples in the dataset. The resulting ratio represents the proportion of positive samples that the model correctly identified. This metric is important in evaluating the performance of deep learning models, particularly in applications such as image classification, speech recognition, and natural language processing. In other words, it's the ratio between the positive classified samples correctly and the total number of positive samples.

$$Recall = \frac{TP}{TP + FP} \tag{2}$$

Equation (2) is the recall representation in terms of true positives and false negatives.

3.2.5. Precision-recall curve

If the model has high precision and recall, it knows how to predict the sample as positive and most positive and does not miss or expect them as negative.

Bishop and Nasrabadi [28] elucidates that the Precision-Recall curve performs as a visual definition of how precision and recall change as the classification threshold undergoes variations. This curve essentially illustrates how a classifier performs across various operational settings. Additionally, the author underlines that the manipulation of the classification threshold allows for the control of the trade-off between precision and recall. Specifically, lowering the threshold tends to increase recall but may decrease precision and vice versa.

3.2.6. F1-Score

We can use F1-Score to combine Recall and Precision into a single metric. If the algorithm is getting a high score for F1, it means that Precision and Recall are also high values.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * (TP + FP + FN)} \tag{3}$$

Equation (3) is the F1-score representation in terms of precision and recall.

In addition, [30] mentions that F1-Score is used to assess how well a system, like a machine, performs compared to a human evaluator in detecting errors. It's calculated as the harmonic mean of Precision and Recall based on the number of errors both the computer and human evaluator identify using the equation above.

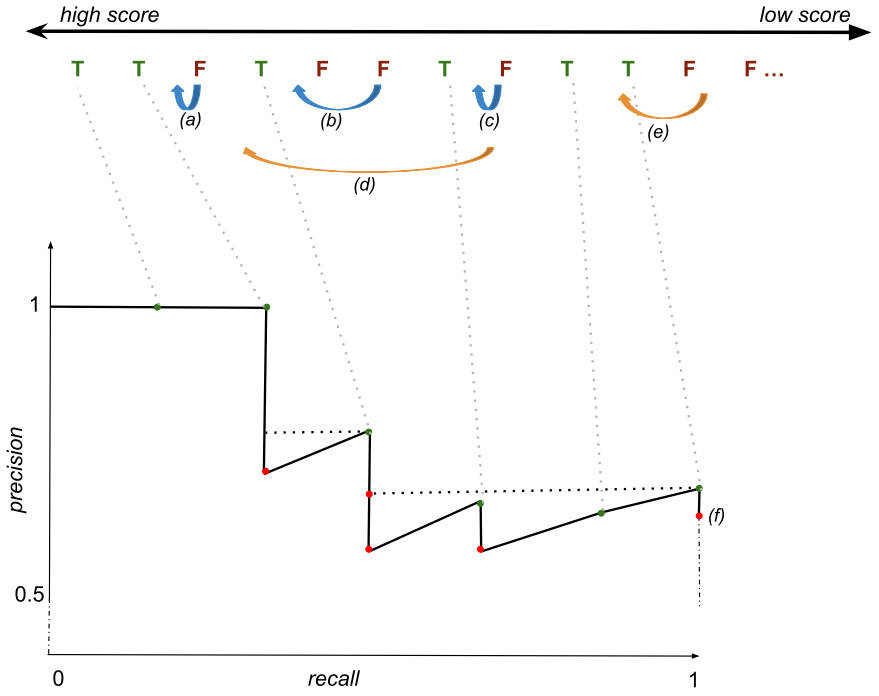


Fig. 5. Precision x Recall Curve, and mAP computation.

3.3. Mean average precision (mAP)

For calculating the mAP, it is taken the mean, over all the APs, of the interpolated AP for each class, as shown in Fig. 5 [31].

$$mAP = \frac{1}{|nclasses|} \sum_{ceclasses} \frac{|TP_c|}{|FP_c| + |TP_c|} \tag{4}$$

Equation (4) is the mAP representation in terms of true positives and false positives.

Fig. 5 illustrates the Precision-Recall curve for a specific object class with six ground truth instances. The black curve is a representation of the precision and recall values computed from a sequence of true-positive (TP) and false-positive (FP) detections ordered by score (top). The dashed line above the solid line area is the outcome of replacing each precision with the maximum at the same or higher recall. The Average Precision is the total area enclosed by the solid and dashed lines.

The arrows labeled (a-e) in the figure highlight the effect of positive perturbations on the FP detection scores. Blue arrows (a-c) indicate perturbations that have no impact on the AP: (a) the order of detections does not change; (b) the detection swaps places with another FP; (c) the detection swaps places with a TP, but a higher-recall TP (f) has higher precision, so there is no change to the area under the filled-in curve (pink shading). On the other hand, orange arrows (d–e) illustrate perturbations that do affect AP: (d) the same FP as (c) is moved beyond a TP that does appear on (hence affect) the filled-in curve; (e) the FP moves past a single TP, altering the filled-in curve as far away as 0.5 recall." [31].

3.3.1. Confusion matrix

It is a graph metric for visualization of the classes. This helps us evaluate/assess our model in predicting each category and the miss classifications between the types in our model. In Fig. 6 (modified from [32]), we can see how the values are distributed in the matrix. Yet, Fig. 7 (modified from [32]) shows an example of diagnosing cancer and no cancer (negative).

Moreover, in [33] is an example of how a confusion matrix can be helpful for interpreting some results of Deep Learning. In their work is proposed research, in which they try to find out the categories that are getting mixed up in classification. This method can also help to notice if those categories have a lot of similarities. For instance, the author used it to group two specific classes together only when the algorithm could tell them apart well. Conversely, when the classifier confuses classes, it means those classes are similar and hard for the classifier to distinguish properly.

3.4. Architecture and topology implemented

The summary of the topology for this work is shown in Fig. 8. The first step was to collect the data. For the private dataset, we cropped and merged it with the other datasets on the Roboflow website. After that, we did the pre-processing (augmentation). Then, we fed the data into our modified Mask RCNN. At the end of these steps, we got the images with the segmented area of the cells and their respective classes.

		Predicted condition	
	Total 8+4=12	Cancer 7	Non-cancer 5
Actual condition	Cancer 8	6	2
	Non-cancer 4	1	3

Fig. 6. Confusion matrix layout.

		Predicted condition	
	Total population = P+N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Fig. 7. Confusion matrix cancer and negative cancer.

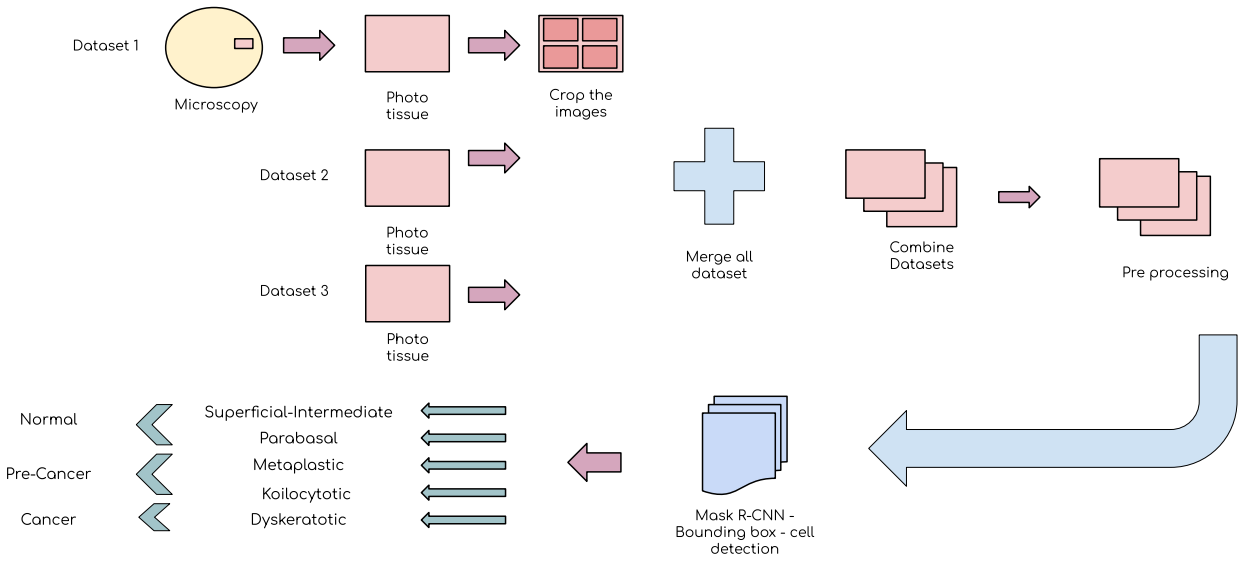


Fig. 8. Whole Architecture/Topology.

For this research, Mask RCNN was employed due to its effectiveness in classification, detection, and segmentation capabilities. Yet, because the annotations are made of polygonal geometries and the cells have different shapes, it can be easier for the network to distinguish between the classes and then improve the precision (mAP). In addition, Mask RCNN considers the segmentation area pixel-to-pixel loss when processing its output total loss. This metric is a measurement that tells us many important parameters when we are training a deep learning algorithm, such as the training error (training error to be minimized) and overfitting/underfitting. Choosing an inadequate loss function for a specific application/network can bring poor results.

In this scenario, Mask RCNN has three different losses, $L = L_{class} + L_{bbox} + L_{mask}$, the loss for the classification (classes), the loss for the bounding box, and the loss for the mask (segmentation), respectively [34]. The first two losses were implemented in the Faster RCNN (previous Mask RCNN version), L_{class}: a discrete probability distribution over K + 1 categories, it also includes the RPN (Regional Proposal Network) classification loss; and L_{bbox}: bounding-box regression offsets [35], including as well the RPN bounding box. However, in contrast to the Faster RCNN, the Mask RCNN included in the L_{mask}, and it allows the network to generate masks for every class without competition among classes [34].

Table 7
Mask RCNN ResNeXt configurations.

GPU_COUNT	1
IMAGES_PER_GPU	1
Batch size = GPUs * images/GPU	1*1 = 1
STEPS_PER_EPOCH	220
VALIDATION_STEPS	200
TRAIN_ROIS_PER_IMAGE	512
MAX_GT_INSTANCES	256
POST_NMS_ROIS_INFERENCE	2000
POST_NMS_ROIS_TRAINING	2000
DETECTION_MAX_INSTANCES	400
DETECTION_MIN_CONFIDENCE	0.5

Table 8
Augmentation Experiments 5 classes ResNeXt 140 layers.

Augmentation	mAP	mAR	F1-score
940 train images	0.34	0.79	0.47
1900 train images	**0.48**	**0.80**	**0.60**
2800 train images	0.35	0.58	0.44

Table 9
Augmentation experiments 5 classes ResNeXt 143 layers.

Augmentation	mAP	mAR	F1-score	Overfitting
No aug	0.22	0.55	0.32	Yes
2800 test images	**0.599**	**0.86**	**0.701**	**No**

4. Experiments

Experimental results are presented in this section. In order to realize the experiments, a machine was used, and it has the following specifications: Intel® Core™ i7-8700 K CPU @ 3.70 GHz × 12 NVIDIA GeForce RTX 2080 Ti/PCIe/SSE2. Moreover, the basic Mask RCNN configuration is presented in Table 7.

4.1. Pre-processing and augmentation

During the initial experiments using Mask RCNN ResNeXt for the dataset SIPaKMeD, it was realized that the loss graph was over-fitting. Therefore, modifications were needed to overcome this issue. In order to overcome this issue, all the data sets were combined. To increase the size of the dataset, augmentation was performed. For instance, rotation (between -15° and +15°), shear (±15° horizontal, ±15° vertical), and mosaic were applied to the images.

Table 8 contains the experiments done with the ResNeXt 140 layers, i.e., the ResNeXt with convolution block 3x4 instead of 3x8. The augmentation of 2800 images in the training dataset leads to the best statistical matrices.

In one set of experiments, ResNeXt with 143 layers was used. The first experiment was conducted without augmentation, while the second experiment involved training with 2800 augmented images.

In the absence of augmentation, the first experiment exhibited signs of overfitting. This conclusion is drawn from the loss graph, specifically looking at Fig. 9. The primary contributors to this overfitting were the losses associated with bounding box prediction and the RPN's bounding box loss (represented by the red rectangle).

The RPN consists of two main components. The first component is the classification loss, which evaluates the RPN's ability to distinguish between foreground (object) and background regions. The second component is the bounding box regression loss, which measures the discrepancy between the predicted bounding box coordinates (e.g., width, height, x, and y offsets) and the ground truth bounding box coordinates for positive anchor boxes.

On the other hand, in the second experiment where augmentation was applied (using 2800 training images), the issues observed in the total bounding box losses and the RPN related to the bounding box (both highlighted in the red rectangle in Fig. 10) were soothed. We need to remember that the total loss function consists of the sum of the three elements (mask, box, and classes), including the RPN ones. Enhancing any one of these components will lead to an improvement in the overall loss, in this case, we want to address the overfitting and decrease the total loss). Yet, Table 9 provides the statistical parameters related to these improvements.

4.2. New mask RCNN backbone

In order to improve the performance, we changed the layer of the Resnet backbone (more information is provided in the Data Availability section). After running a couple of experiments, we realized that by increasing the number of layers, the results have improved, see Table 10.

In our initial experiment, we found out that the precision achieved by Resnet 101 was not insufficient, so we changed the backbone to ResNeXt 152 layers. Yet, ResNeXt 152 is demanding regarding operational complexity. Because we didn't want to downgrade other

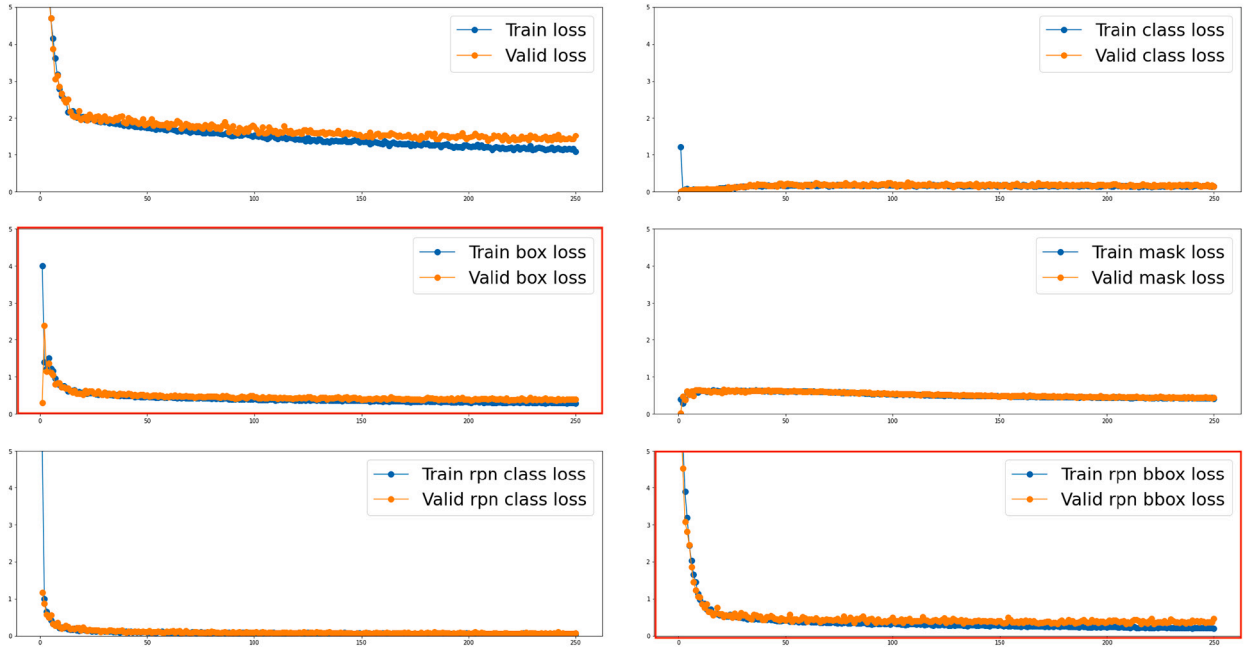


Fig. 9. ResNeXt 143 layers no augmentation and overfitting.

Table 10
Backbone's comparison 5-class.

Backbone	mAP	mAR	F1-score	α (used for train)
Resnet 50	0.49	0.81	0.61	0.001
Resnet 101	0.56	0.83	0.67	0.001
ResNeXt 50	0.52	0.80	0.63	0.001
ResNeXt 101	0.47	0.76	0.58	0.001
ResNeXt 140	0.52	0.78	0.63	0.001
ResNeXt 143	**0.6**	**0.86**	**0.701**	**0.001**
ResNeXt 152	loss Nan	loss Nan	loss Nan	0.0001

parameters involving the image size and quality. For example, we need to train the network using $\alpha = 0.00001$, which would take around one week to train. To solve the complexity problem and not lose the precision, the 152 ResNeXt was replaced with 143 layers (more information is provided in the Data Availability section).

The basic structure of the Resnet network is five different convolution block layers (more information is provided in the Data Availability section). For ResNeXt 143, we downgraded ResNeXt 152, the third convolution block, to 4 times instead of 8 times and added one more time to the last convolution block. In comparison to ResNeXt 101-layer, results have significantly improved.

4.3. Five classes experiments

Running the experiments for five classes, we reached an mAP of 59.9%, about 60%, mAR of 86%, and F1-Score of 70%. The loss function is shown in Fig. 10. We have overcome the over-fitting problem, and the network performs learning. Moreover, looking at Fig. 11, the most mistaken between classes is Meta (Metaplastic) with Koil (Koilocytotic); both are pre-cancer parameters (categories), and the BG is the background class, which means the cells found weren't seen before in the annotations file. Fig. 12 is one example of the output images from our architecture.

4.4. Three classes experiments

Running the experiments for three classes, we achieved an mAP of 59.2%, mAR of 89%, and F1-score of 71.14%. Fig. 13 shows the loss functions (and it is not over-fitting). Moreover, looking at Fig. 14, the most mistaken between classes is Pre-cancer with normal cells. Yet, Fig. 15 is one example of the output image from our architecture.

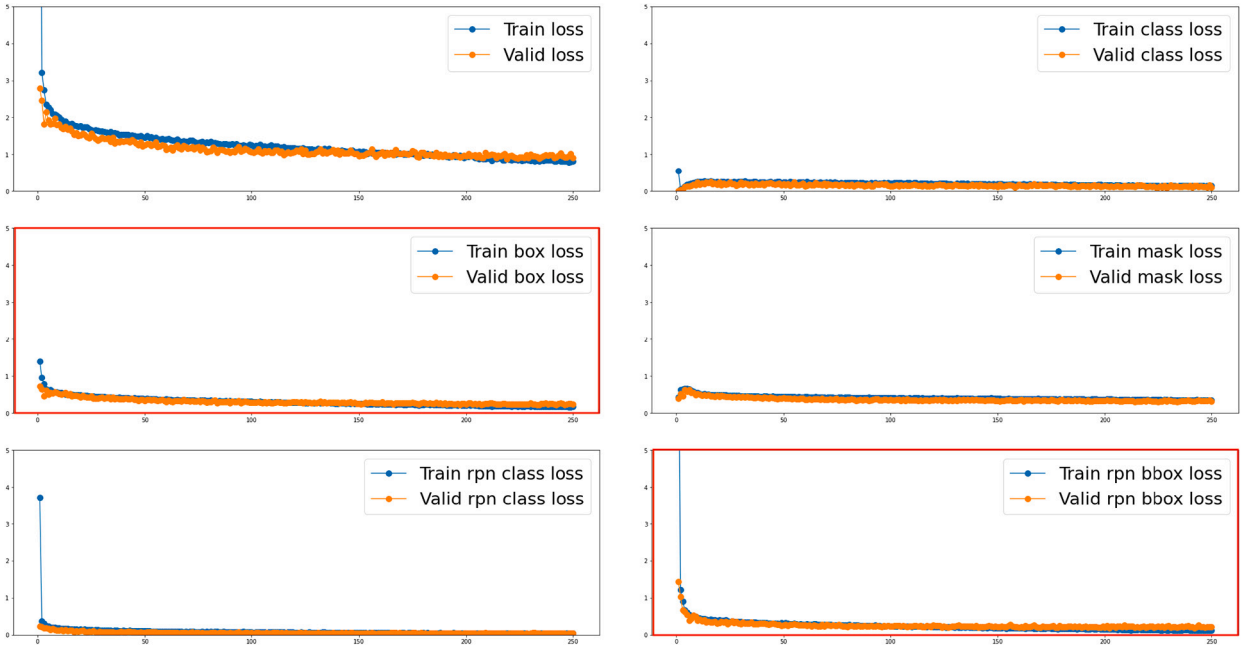


Fig. 10. Loss 5-class; No overfitting and underfitting in all the losses.

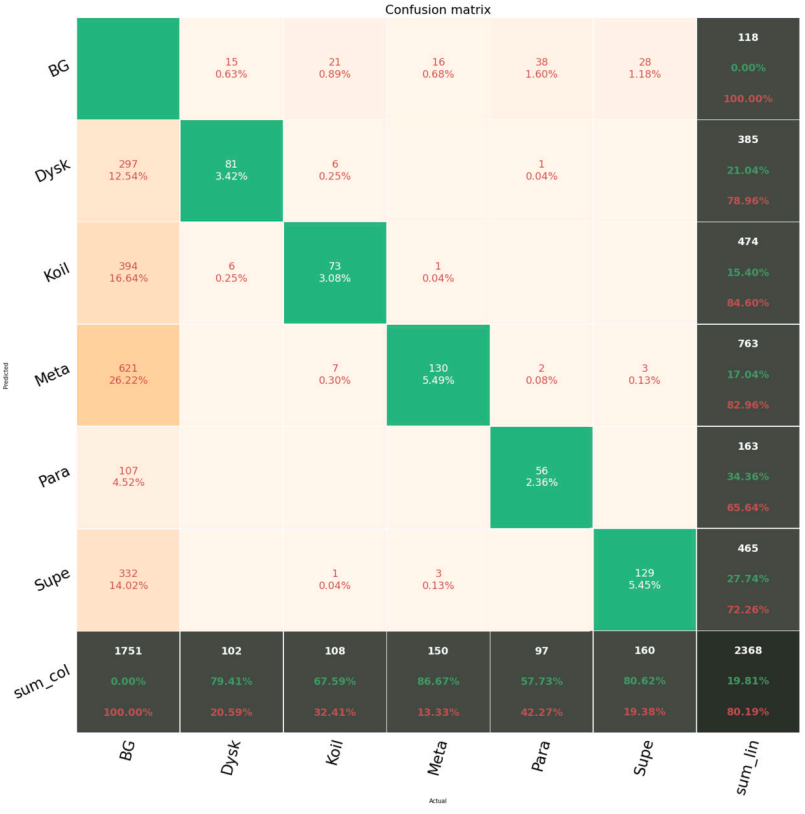


Fig. 11. Confusion matrix 5-class.

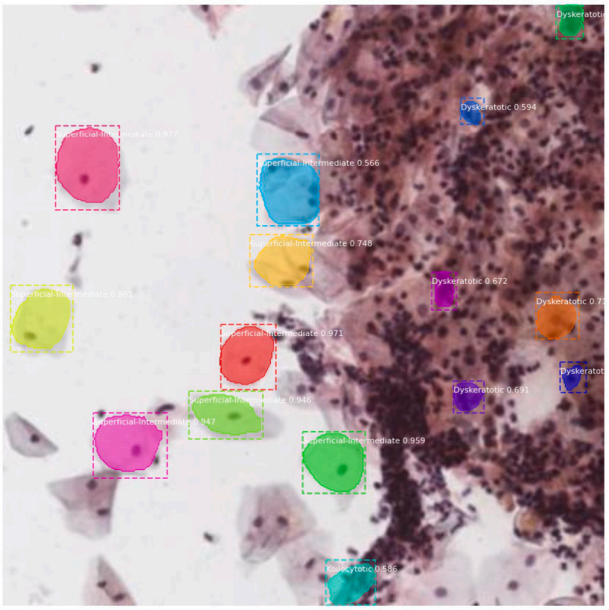


Fig. 12. Example 5-class segmentation, classification, and detection.

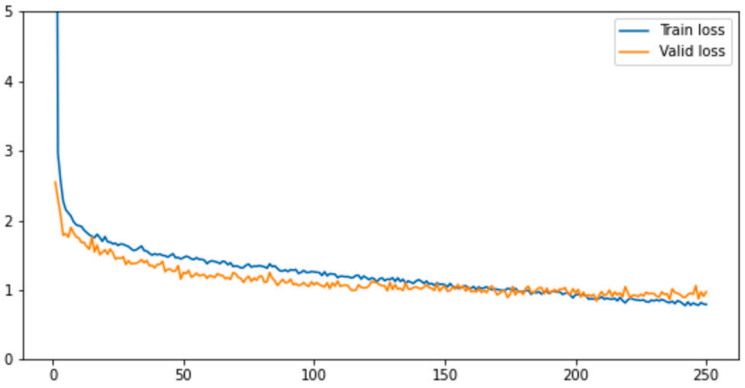


Fig. 13. Loss 3-class; No overfitting and underfitting.

Table 11
Ablation Test 2rd Conv layer.

Number of layers	mAP	mAR	F1-score
1 layer	0.236	0.641	0.346
2 layers	0.2315	0.57	0.33
3 layers	**0.31**	**0.69**	**0.42**
4 layers	0.312	0.618	0.415
5 layers	Nan	Nan	Nan

4.5. Ablation test

4.5.1. Ablation for convolution layers

The purpose of the ablation test in Deep Learning is to evaluate the impact of individual features on the overall system. This involves excluding such a feature or a layer during training and testing to assess the results. From this perspective, we performed some experiments and analyzed the results obtained from them.

The ablation tests for the second, third, and last convolution layers in the ResNeXt for 5-class are presented in Tables 11, 12, and 13 respectively, using learning rate (α) as 0.0001, and 250 epochs. We specifically chose to conduct experiments on these convolution layers, as the first convolution layer follows the same pattern across all Resnet and Resnext models, consisting of just one layer. We did not perform the ablation test on the fourth layer due to the large number of convolution layers, which we maintained at 36.

Fig. 14. Confusion matrix 3-class.

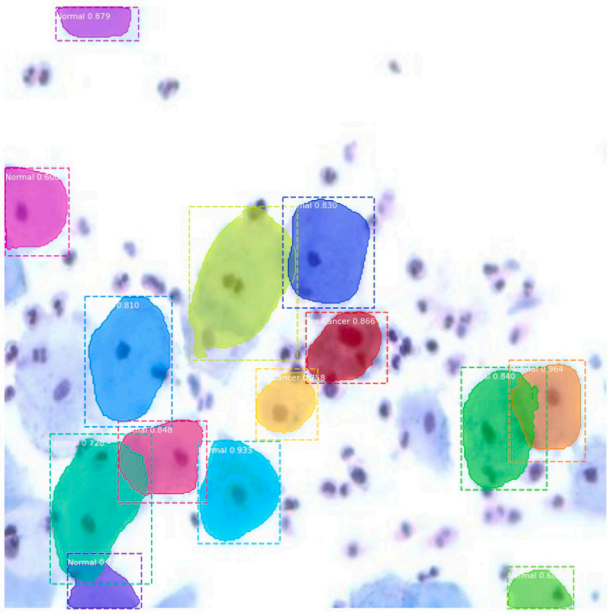


Fig. 15. Example 3-class segmentation, classification, and detection.

Table 12
Ablation Test 3rd Conv layer.

Number of layers	mAP	mAR	F1-score
1 layer	0.29	0.59	0.39
2 layers	0.266	0.64	0.37
3 layers	0.52	0.78	0.63
4 layers	0.31	0.69	0.42
5 layers	**0.36**	**0.66**	**0.46**

Table 13
Ablation Test 5th Conv layer.

Number of layers	mAP	mAR	F1-score
1 layer	0.27	0.67	0.38
2 layers	0.28	0.66	0.4
3 layers	0.27	0.69	0.39
4 layers	**0.31**	**0.69**	**0.42**
5 layers	Nan	Nan	Nan

Table 14
Ablation Test 1st Activation Layer.

Activation Method	mAP	mAR	F1-score
Relu	**0.6**	**0.86**	**0.701**
Softmax	0.4	0.68	0.5
Sigmoid	0.49	0.82	0.41
Elu	Nan	Nan	Nan
Selu	Nan	Nan	Nan

The results indicated that increasing the number of layers doesn't always correspond at the end to an increase in precision. The Nan means that the computer was not computationally capable of running using the settings. Besides that, the best performance was gotten using 5 layers for the 3rd convolution layer, and however, when we use $\alpha = 0.001$ in the 3rd Conv layer with 5 layers, the results also get Nan.

4.5.2. Ablation for activation layer

Activation layers enable neural networks to capture complex, non-linear relationships between inputs and outputs by applying a mathematical function to the network's outputs. This non-linear transformation empowers the neural network to learn and represent intricate patterns and relationships in the data. For this ablation test, we changed the activation layer to use the activation methods presented in Table 14. Some methods were able to run utilizing the current settings, and *Relu* still being the one that achieved higher statistical parameters.

5. Results and discussion

For discussion of our work, we evaluate the results, comparing them to the doctor's point of view and analyzing our results. Moreover, we also made a comparison with the research in this field that also used Deep Learning for segmentation and detection of the cells and cancer stages.

5.1. Doctor's evaluation and comparison

For this comparison, we shared the test dataset and the output from the architecture with a Medical Consultant to analyze 30 images for the 3 and 5-class experiments. The Medical Consultant also completed Table 15 (5-class), Table 16 (3-class) with the information of how many cells were classified incorrectly by the algorithm for each class.

For the 5-class, the overall average was 76.51% of accuracy; however, the algorithm could predict most cases of Dyskeratotic (that is, the cancer cell), and the most mistaken classes were between Metaplastic and Koikocytic that both belong to the pre-cancer group. For the 3-class, the total average was higher, achieving 84.31% accuracy, and the class with fewer incorrect cells was normal, followed by cancer and pre-cancer. In conclusion, of the 30 images the doctor analyzed, the pre-cancer cells were the most difficult to predict.

Moreover, we also provide a report with the pre-cancer (dangerous cells) and cancer (carcinogen cells) and their respective locations. The provided location is determined by the bounding box, where the given X and Y coordinates represent the center of the rectangle. In simpler terms, they indicate the center of the cell. We can see an example of a report for two different images in Fig. 16.

Table 15
Table for comparison 5-class.

Image	Number of cells	Number of correct cells	Superficial-Intermediate incorrect	Parabasal incorrect	Metaplastic incorrect	Koilocytotic incorrect	Dyskeratotic incorrect
Image 1	24	22	0	0	1	1	0
Image 2	19	14	1	0	0	4	0
Image 3	23	20	0	0	1	2	0
Image 4	12	10	0	0	0	2	0
Image 5	25	24	0	0	0	1	0
Image 6	26	25	0	0	0	1	0
Image 7	22	19	0	0	0	3	0
Image 8	31	30	0	0	0	1	0
Image 9	32	32	0	0	0	0	0
Image 10	25	15	0	0	3	7	0
Image 11	14	12	0	0	2	0	0
Image 12	21	11	0	0	4	6	0
Image 13	13	10	0	0	3	0	0
Image 14	14	13	0	0	0	1	0
Image 15	12	6	0	0	0	5	1
Image 16	7	5	0	0	1	1	0
Image 17	12	8	0	0	2	2	0
Image 18	14	2	0	0	2	0	0
Image 19	17	14	0	0	3	0	0
Image 20	25	21	0	0	4	0	0
Image 21	21	17	0	0	1	3	0
Image 22	21	16	0	0	4	1	0
Image 23	27	15	0	0	1	1	0
Image 24	20	16	0	2	2	0	0
Image 25	16	0	0	0	0	0	0
Image 26	13	9	1	1	0	1	1
Image 27	21	14	0	0	2	5	0
Image 28	15	11	0	0	2	2	0
Image 29	16	15	0	0	0	1	0
Image 30	21	17	0	0	3	1	0
Total	579	443					
Percentage	76.51122625						

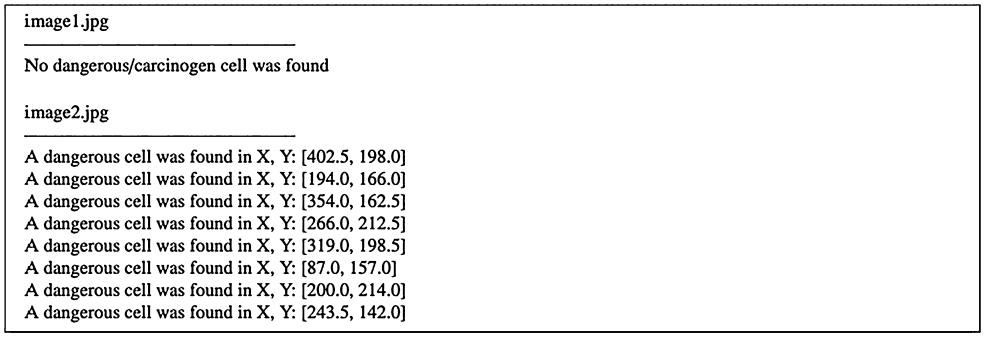


Fig. 16. Report Example.

5.2. Research comparison

In this comparison, we conducted research using papers that focused on complete tissue cell segmentation and classification. To evaluate the performance, we employed the mAP statistical method, as shown in Table 17. Besides that, in Table 18, we provide the dataset information for the dataset details.

One of the papers we considered was [23], which demonstrated a low mAP when using the same benchmark dataset that we utilized in our study. Additionally, in [24], a better mAP was achieved for the 10-class category using a different dataset with more categories. This suggests that increasing the number of classes can sometimes aid the network in learning specific features and effectively distinguishing between different cancer levels and normal cells. Moreover, it is worth noting the variation in sample sizes employed in each of the papers; the greater the amount of data available, the higher the level of precision that can be achieved.

Another article we included in this comparison was [36], which presented an end application for determining whether a cell is carcinogenic or not. However, it did not provide additional information regarding cancer classification, such as the degree or location. In our study, we obtained similar mAP results by employing different datasets, methods (modified Mask RCNN), and

Table 16
Table for comparison 3-class.

Image	Number of cells	Number of correct cells	Normal incorrect	Pre-cancer incorrect	Cancer incorrect
Image 1	31	26	1	4	0
Image 2	21	17	1	3	0
Image 3	26	23	0	3	0
Image 4	16	14	1	1	0
Image 5	45	44	0	1	0
Image 6	39	35	1	3	0
Image 7	28	23	0	5	0
Image 8	33	31	1	1	0
Image 9	14	13	0	1	0
Image 10	24	22	0	0	2
Image 11	40	38	1	1	0
Image 12	7	7	0	0	0
Image 13	7	7	0	0	0
Image 14	28	21	0	7	0
Image 15	30	27	0	3	0
Image 16	22	19	0	3	0
Image 17	21	19	1	1	0
Image 18	21	20	0	1	0
Image 19	12	6	0	1	5
Image 20	22	13	0	0	9
Image 21	16	14	0	2	0
Image 22	12	8	0	0	4
Image 23	20	19	0	0	1
Image 24	6	5	0	1	0
Image 25	21	19	0	2	0
Image 26	11	9	0	2	0
Image 27	32	27	0	5	0
Image 28	26	23	0	3	0
Image 29	11	8	0	3	0
Image 30	21	2	0	2	0
Total	663	559			
Percentage	84.31372549				

Table 17
Deep Learning Classification and Segmentation Comparison.

Name	Dataset	Algorithm	Detection/Classification/Segmentation	Results	Tool Cell location
[37]	Herlev and Private	Improved YOLOv3	Detection and Classification	mAP of 78.87%	No
[36]	Harlev and Private	Trainable Weka Segmentation	Detection	Acc of 98.88%	Yes
[23]	SIPaKMeD	Faster R-CNN	All	mAP of 0.37798, and AR of 0.64 (5-class)	No
[24]	Herlve and private	YOLOv3	All	mAP of 0.6 (10-class)	No
Our model	SIPaKMeD and private	Modified Mask-RCNN	All	mAP 0f 0.6 and mAR of 0.86 (5-class)	Yes

Table 18
Datasets information.

Name	Dataset	Number of classes	Number of images and cells	Name of the classes
[37]	Herlev and Private	7 classes	54,000 cells	LSIL, HSIL, SCC, AEC, AIS, AGC, and EA
[36]	Harlev and Private	2 classes	917 single cells and 557 full slides	Abnormal, and Normal
[23]	SIPaKMeD	5 classes	966 image and 4,490 cells	Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate
[24]	Herlve and private	10 classes	12,909 images and 58,995	Normal, ACUS, ASCH, LSIL, HSIL, AGC, ADE, VAG, MON, and DYS
Our model	SIPaKMeD and private	3 and 5 classes	1342 images and about 5 thousand cells	Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate

classes. Furthermore, we concluded our research by offering a perspective from the doctor's point of view and providing a report for quick clinic analysis.

Considering the following notations for Table 18: Adenocarcinoma (ADE), Vaginalis trichomoniasis (VAG), Monilia (MON), dysbacteriosis (DYS), LSIL (low-grade squamous intraepithelial lesion cells), HSIL (High squamous epithelial cells), SCC (Squamous carcinoma cells), AEC (Cervical gland cells), AIS (Cervical adenocarcinoma in situ), AGC (Cervical canal adenocarcinoma), and EA (Endometrial adenocarcinoma).

6. Conclusion and future work

Cervical cancer has affected women all over the world. Early diagnosis can save the lives of many women. However, women have to wait longer before being seen by consultants. Hence, there is an urgent need for reliable and fast detection and diagnosis techniques. Artificial intelligence techniques such as deep NN have great potential in this field to automate the process of detection and diagnosis.

This work implements a novel Deep Learning architecture based on a mask region-based Convolutional neural network for cervical cancer classification and segmentation using tissue cells. The proposed architecture can be employed for any type of data set and application in this field without adding many new layers (adding new layers makes the system computationally very cumbersome). With the help of our novel architecture, higher values of mAP of 60% and F1-score of 70% (5-class), mAP of 59.2%, and F1-score of 71.14% (3-class) were achieved in comparison to the previous work.

Our proposed system can deliver a readable report, enabling the medical consultant to identify the malicious cells and cancer stages in a short time. In the future, we want to increase the precision for all the classes, especially for the pre-cancer (usually the poorest evaluated), and expand this work to be more user-friendly and to be made available to health systems worldwide.

CRediT authorship contribution statement

Carolina Rutili de Lima: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Said G. Khan:** Conceptualization, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Syed H. Shah:** Conceptualization, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Luthiari Ferri:** Conceptualization, Data curation, Formal analysis.

Declaration of competing interest

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

Data availability

The data, the code, and support information used for this research are available at the following link: here.

Review and/or approval by an ethics committee was not needed for this study because the data supplied for this study does not contain any sensitive information, such as names or ages, so no ethical disclosure is required.

References

[1] WHO, Cervical cancer, https://www.who.int/news-room/fact-sheets/detail/cervical-cancer/, 2022. (Accessed 10 July 2022), Online.
[2] d.P. Sociedade Brasileira, Patologista – O profissional do diagnóstico, https://www.sbp.org.br/patologista-o-profissional-do-diagnostico/, 2016. (Accessed 12 July 2022), Online.
[3] M.S. Khan, F.Y. Raja, G. Ishfaq, F. Tahir, F. Subhan, B.M. Kazi, K.A. Karamat, Pap smear screening for pre-cancerous conditions of the cervical cancer, Pak. J. Med. Res. 44 (3) (2005) 111–113.
[4] B.E. Sirovich, H.G. Welch, The frequency of pap smear screening in the United States, J. Gen. Intern. Med. 19 (3) (2004) 243–250.
[5] M.d. Saúde Brasil, Papanicolau (exame preventivo de colo de útero), https://bvsms.saude.gov.br/papanicolau-exame-preventivo-de-colo-de-utero/, 2011. (Accessed 12 July 2022), Online.
[6] D. Lee, S.N. Yoon, Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges, Int. J. Environ. Res. Public Health 18 (1) (2021) 271.
[7] M. Koutinas, A. Kiparissides, E.N. Pistikopoulos, A. Mantalaris, Bioprocess systems engineering: transferring traditional process engineering principles to industrial biotechnology, Comput. Struct. Biotechnol. J. 3 (4) (2012) e201210022.
[8] F. Sauer, A. Fritsch, S. Grosser, S. Pawlizak, T. Kießling, M. Reiss-Zimmermann, M. Shahryari, W.C. Müller, K.-T. Hoffmann, J.A. Käs, et al., Whole tissue and single cell mechanics are correlated in human brain tumors, Soft Matter 17 (47) (2021) 10744–10752.
[9] A. Ghoneim, G. Muhammad, M.S. Hossain, Cervical cancer classification using convolutional neural networks and extreme learning machines, Future Gener. Comput. Syst. 102 (2020) 643–649.
[10] M.I. Waly, M.Y. Sikkandar, M.A. Aboamer, S. Kadry, O. Thinnukool, Optimal deep convolution neural network for cervical cancer diagnosis model, Comput. Mater. Continua 70 (2) (2022) 3297–3309.
[11] K.P. Win, Y. Kitjaidure, K. Hamamoto, T. Myo Aung, Computer-assisted screening for cervical cancer using digital image processing of pap smear images, Appl. Sci. 10 (5) (2020) 1800.
[12] A.-u. Rehman, N. Ali, I. Taj, M. Sajid, K.S. Karimov, et al., An automatic mass screening system for cervical cancer detection based on convolutional neural network, Math. Probl. Eng. 2020 (2020).
[13] K. Sabeena, C. Gopakumar, A hybrid model for efficient cervical cell classification, Biomed. Signal Process. Control 72 (2022) 103288.
[14] A.D. Jia, B.Z. Li, C.C. Zhang, Detection of cervical cancer cells based on strong feature cnn-svm network, Neurocomputing 411 (2020) 112–127.
[15] O. Yaman, T. Tuncer, Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images, Biomed. Signal Process. Control 73 (2022) 103428.
[16] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, R. Sarkar, A fuzzy rank-based ensemble of cnn models for classification of cervical cytology, Sci. Rep. 11 (1) (2021) 1–18.
[17] R. Pramanik, M. Biswas, S. Sen, L.A. de Souza Júnior, J.P. Papa, R. Sarkar, A fuzzy distance-based ensemble of deep models for cervical cancer detection, Comput. Methods Programs Biomed. 219 (2022) 106776.

[18] E. Hussain, L.B. Mahanta, C.R. Das, R.K. Talukdar, 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 Cell 65 (2020) 101347.
[19] H. Chen, J. Liu, Q.-M. Wen, Z.-Q. Zuo, J.-S. Liu, J. Feng, B.-C. Pang, D. Xiao, Cytobrain: cervical cancer screening system based on deep learning technology, J. Comput. Sci. Technol. 36 (2) (2021) 347–360.
[20] S. Cheng, S. Liu, J. Yu, G. Rao, Y. Xiao, W. Han, W. Zhu, X. Lv, N. Li, J. Cai, et al., Robust whole slide image analysis for cervical cancer screening using deep learning, Nat. Commun. 12 (1) (2021) 1–10.
[21] X. Tan, K. Li, J. Zhang, W. Wang, B. Wu, J. Wu, X. Li, X. Huang, Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study, Cancer Cell Int. 21 (1) (2021) 1–10.
[22] K.H.S. Allehaibi, L.E. Nugroho, L. Lazuardi, A.S. Prabuwono, T. Mantoro, et al., Segmentation and classification of cervical cells using deep learning, IEEE Access 7 (2019) 116925–116941.
[23] A.F. Sampaio, L. Rosado, M.J.M. Vasconcelos, Towards the mobile detection of cervical lesions: a region-based approach for the analysis of microscopic images, IEEE Access 9 (2021) 152188–152205.
[24] Y. Xiang, W. Sun, C. Pan, M. Yan, Z. Yin, Y. Liang, A novel automation-assisted cervical cancer reading method based on convolutional neural network, Biocybern. Biomed. Eng. 40 (2) (2020) 611–623.
[25] M.E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, A. Charchanti, Sipakmed: a new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 3144–3148.
[26] E. Hussain, Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer, Data Brief (2019).
[27] A. Sharma, Mean average precision (mAP) using the COCO evaluator, https://pyimagesearch.com/2022/05/02/mean-average-precision-map-using-the-coco-evaluator/, 2022. (Accessed 20 October 2022), Online.
[28] C.M. Bishop, N.M. Nasrabadi, Pattern Recognition and Machine Learning, vol. 4, Springer, 2006.
[29] H. Schütze, C.D. Manning, P. Raghavan, Introduction to Information Retrieval, vol. 39, Cambridge University Press, Cambridge, 2008.
[30] H. Huang, H. Xu, X. Wang, W. Silamu, Maximum f1-score discriminative training criterion for automatic mispronunciation detection, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (4) (2015) 787–797.
[31] P. Henderson, V. Ferrari, End-to-end training of object class detectors for mean average precision, in: Asian Conference on Computer Vision, Springer, 2016, pp. 198–213.
[32] Wikipedia, Confusion matrix, https://en.wikipedia.org/wiki/Confusion_matrix, 2022. (Accessed 25 October 2022), Online.
[33] R. Susmaga, Confusion matrix visualization, in: Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM '04 Conference Held in Zakopane, Poland, May 17–20, 200, Springer, 2004, pp. 107–116.
[34] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[35] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[36] W. William, A. Ware, A.H. Basaza-Ejiri, J. Obungoloch, A pap-smear analysis tool (pat) for detection of cervical cancer from pap-smear images, Biomed. Eng. Online 18 (2019) 1–22.
[37] D. Jia, Z. He, C. Zhang, W. Yin, N. Wu, Z. Li, Detection of cervical cancer cells in complex situation based on improved yolov3 network, Multimed. Tools Appl. 81 (6) (2022) 8939–8961.