

<https://dx.doi.org/10.17488/RMIB.46.2.1462>

E-LOCATION ID: 1462

## Pix2Pix Generative Adversarial Network for Cellular Nuclei and Cytoplasm Segmentation on Pap Smear Images

### Red Generativa Antagónica Pix2Pix para la Segmentación de Núcleos Celulares y Citoplasma en Imágenes de Frotis de Papanicolaou

Francisco Javier Castro Cortés<sup>1</sup> , Carlos Eric Galván-Tejada<sup>1</sup>  , Erika Acosta Cruz<sup>1</sup> ,  
José M. Celaya-Padilla<sup>1</sup> 

<sup>1</sup>Universidad Autónoma de Zacatecas, Zacatecas - México

<sup>2</sup>Universidad Autónoma de Coahuila, Coahuila - México

#### ABSTRACT

In medical imaging for Pap smear tests, accurately identifying regions of interest, such as the nucleus and cytoplasm, remains a critical challenge due to the complex morphology and overlapping structures in cervical cell images. This complexity increases the risk of misidentification, potentially leading to false positives in computer-assisted diagnosis. To address this issue, this study introduces a novel approach by developing and evaluating a framework for the precise segmentation of nuclei and cytoplasm in cervical cell images using a cGAN-based model, Pix2Pix, applied to a dataset validated by specialists. The generated images are compared with target images, converted to binary, and an AND operation is performed to evaluate pixel overlap in the areas of interest. The evaluation metrics highlight a segmentation accuracy of 88.8 % and sensitivity of 89.62 % for nuclei, while for cytoplasm, precision reached 89.62 % and sensitivity 99.34 %. The Jaccard indices were 80.89 % for nuclei and 96.71 % for cytoplasm. These results demonstrate the effectiveness of the model in segmenting nuclei and cytoplasm in cervical cells.

**KEYWORDS:** cancer, cGAN, segmentation, PAP, Pix2Pix

## RESUMEN

En el campo de las imágenes médicas para la prueba de Papanicolaou, identificar con precisión las regiones de interés, como el núcleo y el citoplasma, sigue siendo un desafío crítico debido a la compleja morfología y las estructuras superpuestas en las imágenes de células cervicales. Esta complejidad aumenta el riesgo de identificaciones erróneas, lo que podría llevar a falsos positivos en el diagnóstico asistido por computadora. Para abordar este problema, este estudio presenta un enfoque novedoso mediante el desarrollo y evaluación de un marco para la segmentación precisa de núcleos y citoplasmas en imágenes de células cervicales, utilizando un modelo basado en cGAN, Pix2Pix, aplicado a un conjunto de datos validado por especialistas. Las imágenes generadas se compararon con las imágenes objetivo, se convirtieron a formato binario y se realizó una operación AND para evaluar la superposición de píxeles en las áreas de interés. Las métricas de evaluación destacaron una precisión de segmentación del 88.8 % y una sensibilidad del 89.62 % para los núcleos, mientras que, para el citoplasma, la precisión alcanzó el 89.62 % y la sensibilidad el 99.34 %. Los índices de Jaccard fueron del 80.89 % para los núcleos y del 96.71 % para el citoplasma. Estos resultados demuestran la efectividad del modelo en la segmentación de núcleos y citoplasmas en células cervicales.

**PALABRAS CLAVE:** cáncer, cGAN, segmentación, PAP, Pix2Pix

### Corresponding author

TO: CARLOS ERIC GALVÁN-TEJADA

INSTITUTION: UNIVERSIDAD AUTÓNOMA DE ZACATECAS,  
ZACATECAS - MÉXICO

ADDRESS: CARRETERA ZACATECAS-GUADALAJARA KM. 6, EJIDO  
LA ESCONDIDA, ZACATECAS, C.P. 98160, MÉXICO.

EMAIL: ericgalvan@uaz.edu.mx

### Received:

12 Sep 2024

### Accepted:

11 Mar 2025

### Published:

1 May 2025

## INTRODUCTION

Cervical cancer according to the World Health Organization is a type of cancer that originates in the cells of the cervix, which is the lower part of the uterus that extends into the vagina, and is one of the leading causes of cancer in women worldwide. There are different carcinogenic agents or factors, such as radiation, chemicals, tobacco smoke and infectious agents. Additionally, abnormalities can arise during cell replication. These factors may be hereditary, environmental or related to lifestyle<sup>[1][2][3][4]</sup>.

For diagnosis, the cervical cancer screening method is performed via the Papanicolaou test (PAP) or cervical cytology. Developed by Dr. George Papanicolaou in 1941, it revolutionized the field by allowing early detection of cellular abnormalities in the cervix. This test has become a common practice and a fundamental tool for women's health. The procedure involves collecting a sample of cells from the cervix and nearby areas. In a laboratory, the samples are then examined on smear slides to check for changes in cell morphology and texture. Analyzing these changes is essential for ensuring early diagnosis, as this type of cancer can be detected at very early and opportune stages. <sup>[5][6]</sup>. According to the National Cancer Institute, Cervical Intraepithelial Neoplasia (CIN) is a long phase of preinvasive diseases that proceed to invasive squamous cell cervical cancers, characterized by abnormal changes in the cells of the cervix<sup>[7][8]</sup>. Addressing a complex problem such as the diagnosis of cervical cancer requires technicians and doctors specialized in the field. It also takes time for these professionals to analyze and interpret the images to provide an accurate diagnosis.

Analyzing patterns in abnormal cell growth patterns, provides relevant information for a computer-assisted diagnosis (CAD). Key anatomical features, such as shape, the relationship between nucleus and cytoplasm, color and clustering, among others, are relevant in this type of analysis, since changes in these are indicative of neoplasia<sup>[9][10]</sup>.

In the field of biomedical image segmentation, particularly for PAP images, several studies have explored nuclei and cytoplasm segmentation. Traditional methods not based on Generative Adversarial Networks (GAN) include the Mean-Shift clustering algorithm, which is used to obtain Regions of Interest (ROI) in cellular nucleus segmentation, followed by flexible mathematical morphology to separate overlapping nuclei. Other approaches include the Selective Edge-Enhancement Nuclei Segmentation (SEENS) method, which uses the Canny operator and mathematical morphology to extract edge information and selectively enhance the ROI edges, as well as fuzzy c-means clustering, which distinguishes pixels with similar intensity values located in different spatial regions<sup>[11][12][13][14]</sup>.

GANs are a very powerful tool. Although it is difficult to determine which algorithm performs better than the others, making comparisons is key to help us approach this understanding. A neutral, multifaceted and large-scale study comparing different GAN type models found that hyperparameter optimization and random restarts play a crucial role when comparing these models to determine their best score<sup>[15]</sup>.

Several improvements have been proposed for GAN-based frameworks, including enhanced versions Pix2Pix and U-Net. The models were trained with a dataset of 50 cells images (256x256 resolution), labeled with fluorescence in ROI. These cells were extracted from mouse liver. The Jaccard index was used as an evaluation metric on ROI, highlighting a significant improvement in the proposed method<sup>[16][17]</sup>.

A proposed solution to the problem of insufficient large training datasets in image processing is data augmentation.

This process artificially expands the size of training datasets to prevent overfitting. GAN's can generate artificial images almost indistinguishable from real ones, facilitating the creation of synthetic images that closely mimic the originals. In the medical field, where data acquisition is restricted by various procedures, laws, and prohibitions, including artificial images can significantly enhance the training process. A study demonstrated that training Convolutional Neural Networks (CNN) with artificially augmented datasets increased classification accuracy for cells by up to 12.9 % compared to models trained only with the original HERlev Pap Smear dataset. This approach highlights the importance of synthetic data generation in the medical field to improve the performance of deep learning models<sup>[18]</sup>.

Recently, a study explored synthetic Computed Tomography (CT) for prostate radiotherapy planning using the Pix2Pix model. The main focus was on optimizing the model by making adjustments to the generator, loss function, and hyperparameters, thereby improving the quality of the generated images. The adjusted model was compared with five other models, including U-Net and GAN. The results showed that by using a perceptual loss function and a 9-block ResNet generator, the Mean Absolute Error (MAE) was reduced to the lowest level compared to the other models. Additionally, the highest gamma pass rates were achieved in the CT images<sup>[19]</sup>. Another study introduced HistoGAN, a conditional generative adversarial network (cGAN) designed to generate high-quality synthetic histopathology images for selective augmentation. By incorporating these images selectively into training datasets, the method achieved significant improvements in classification accuracy on two datasets: cervical histopathology (6.7 % increase) and lymph node images (2.8 % increase) ensuring that only high confidence synthetic images were incorporated into training datasets<sup>[20]</sup>.

Several algorithms and frameworks have been explored for the segmentation of nuclei and cytoplasm, each with their own advantages and limitations. Active Contours efficiently detect contours but require significant computational time to minimize energy. K-means clustering efficiently finds solutions but is slow and cannot separate overlapping cells. K-Nearest Neighbor (KNN) provides optimal solutions, is fast, and is noise-tolerant, but its complexity increases as the number of attributes grows. Artificial Neural Networks (ANNs) are also noise-tolerant and can handle multiple instances; however, they are limited by overfitting, high complexity, and significant time consumption. Decision Trees are easy to interpret, but errors during training can lead to inaccurate results. Support Vector Machines (SVMs) offer simple control, but training is slow, and finding optimal parameters for non-linear data can be challenging<sup>[21]</sup>.

In the field of cervical cytology, accurately segmenting the nuclei and cytoplasm in Pap smear images constitutes a significant challenge due to the complex cellular morphology and overlapping structures, which increase the risk of diagnostic errors. The hypothesis posits that a cGAN model trained on a specialist-validated dataset achieves more accurate segmentation compared to traditional methods. To evaluate this hypothesis, a framework is proposed using the SIPaKMeD dataset to create images with color semantic maps. These maps are then used to generate images with segmented nuclei and cytoplasm from various cell types (see Figure 1), and the images are used as training inputs for the Pix2Pix model. After training, a pixel-to-pixel evaluation of the binary test images is performed to assess the number of pixels within each nucleus, yielding true positives and the percentage of correctly classified pixels. Similarly, the pixels that do not belong to the nuclei or cytoplasm are evaluated to determine false positives. The framework includes both quantitative and qualitative evaluations. The evaluation parameters provide a comparison of the model's efficiency in segmenting the five cell types included in the SIPaKMeD dataset<sup>[22]</sup>.

This approach is clinically relevant as it improves the accuracy of CAD in Pap tests. Furthermore, it represents an innovative application of cGAN models in this specific context, highlighting their ability to overcome the limitations of previous techniques and contribute to the early detection of cervical anomalies in a more accurate way. This research addresses an important clinical challenge by proposing a robust framework for segmentation of nuclei and cytoplasm in cervical cell images, aiding early cancer detection and diagnosis. The innovative aspect of the study lies in the integration of a conditional Generative Adversarial Network (cGAN) to achieve high segmentation accuracy, an approach rarely explored in cytological image analysis.

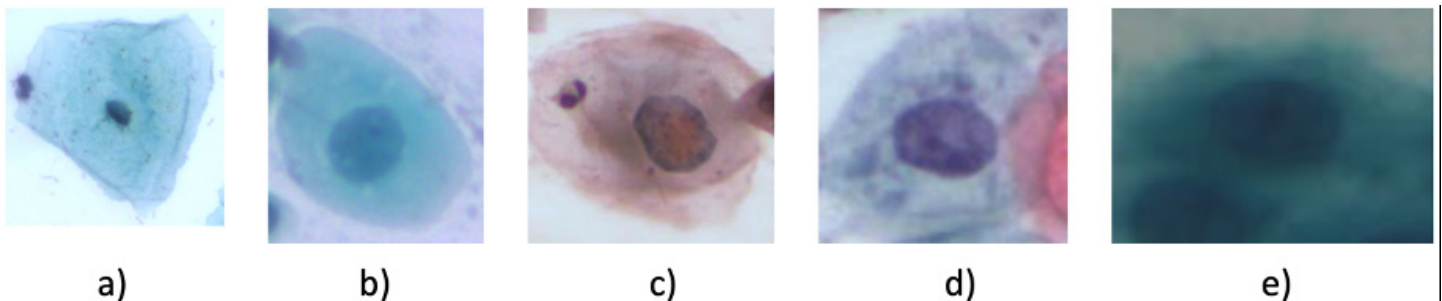
The main contribution in this article is the proposed method for labeling segmented cell nuclei from pap smears using the Pix2Pix cGAN's model. The model is trained with individual images, and a qualitative and quantitative experiment demonstrates the capability of the cGAN in segmenting cell nuclei and cytoplasm.

## MATERIALS AND METHODS

In this section, is detailed the materials and methods that ensure the reproducibility and support the results obtained. It is described the dataset used, including its origin, characteristics, content, and selection, as well as the architecture used and the hyperparameters necessary for the implementation of the cGAN. This part is detailed based on the recommendations and specifications of the Pix2Pix model architecture. Finally, the evaluation of the resulting images is presented under a framework of metric evaluation, allowing a clear and reproducible understanding of the applied analytical methods.

### Dataset description

The SIPaKMeD dataset, which consists of 4049 isolated cell images that have been manually cropped from 966 cell images taken from a Pap smear slide, is widely used in both medical research and in the field of machine vision. There are categories associated with normal cells and cells with the presence of cervical intraepithelial neoplasia. Plissiti *et. al.* categorized according to the position of the cells in the different epithelial layers as well as the degree of maturation they have. The different categories of the dataset are shown in Figure 1 and Table 1, examples of cells from each of the categories are shown.



**FIGURE 1. Imaging cells of five categories: (a) Superficial-Intermediate, (b) Parabasal, (c) Koilocytotic, (d) Dyskeratotic, (e) Metaplastic.**

TABLE 1. Dataset categories.

Category	No. of images	No. of cells
Superficial/Intermediate	126	813
Parabasal	108	787
Koilocytotic	238	825
Metaplastic	271	793
Dyskeratotic	223	813
Total	996	4049

### Architecture model

The architecture of the model used is based on the U-net generating network, taking advantage of the capacity to preserve fine details in the output image. This is where the  $G$  generator is important on this model. As can be seen in Figure 2, the Pix2Pix [23] model is like to the U-Net architecture. However, Skip Connections is incorporated by connecting each of the layers in the encoder and its corresponding in the decoder that's performing activations in between symmetrical layers.

GAN's are artificial neural networks (ANN) which map a random noise vector input ( $z$ ) to an output image ( $y$ ). However, unlike conditional Generative Adversarial Neural Network (cGAN's), they are conditioned by an input image that provides additional information, which in the case of this work is to generate output images (color semantic map) from the original image of a pap smear. It should be noted that cGAN's take better advantage of the morphological characteristics of the cell nuclei[24][25].

The segmented areas of interest, which in this case are the nuclei and cytoplasm in Pap test images, are an important part of identification. These areas will provide relevant information for the training of a model that allows the identification of any type of cervical cancer. To do this, the use of a Deep Learning model called Pix2Pix is proposed. The model is a cGAN, that distinguishes itself by its ability to learn the relationship between an input image and a corresponding output image in a training dataset. This model is based on the mathematical formulation of GANs, where a generating function  $G$  is employed, which transforms an input image  $x$  and a basis distribution function  $F$  into an output image generated  $y$ , while a discriminator function  $D$  evaluates the authenticity of the generated image compared to real images, so given the vector  $x$  and a random vector  $z$  to  $y$ , ( $G:\{z,y\}\rightarrow y$ ), the difference lies in the fact that cGAN's are conditioned to a style depended on the objective, in this case a semantic color map. In summary, this model learns to map input images to target images, such as semantic color maps, by employing a generator that creates images and a discriminator that evaluates their authenticity against real ones.

Mathematically, the objective of the cGAN can be expressed through Equation (1). representing  $x$ , the segmentation mask (segmented cell nucleus and cytoplasm). While  $G(x)$  is the generated image of the segmented core.

$$L_{cGAN}(G, D) = \mathbb{E}_{(x,y)}[\log D(x, y)] + \mathbb{E}_{(x,z)}[\log(1 - D(x, y))] \quad (1)$$

$G$  is the one who makes the attempt to minimize the objective against the antagonist  $D$ , which tries to maximize it. In other words,  $\min G \max D L_{cGAN}(G, D)$ . The definition of the loss of both the generator and the discriminator is given by Equation (2) and (3).

$$L_{\text{gen}}(G, D) = \mathbb{E}_x \left[ \log(1 - D(x, G(x))) \right] + \lambda \mathbb{E}_{(x,y)} [\|y - G(x)\|_1] \quad (2)$$

$$L_{\text{dis}}(G, D) = \mathbb{E}_{(x,y)} [\log(1 - D(x, y))] + \mathbb{E}_x [\log(-D(x, G(x)))] \quad (3)$$

The final goal of the generator and discriminator can be expressed as follows:

$$G^* = \underset{G}{\operatorname{argmin}} L_{\text{gen}}(G, D) = \underset{G}{\operatorname{argmin}} \max_D (L_{\text{cGAN}}(G, D) + \lambda L_{L1}(G)) \quad (4)$$

$$D^* = \underset{D}{\operatorname{argmin}} L_{\text{dis}}(G, D) = \underset{D}{\operatorname{argmin}} \max_G L_{\text{cGAN}}(G, D) \quad (5)$$

The status-of-the-art suggest an architecture of *encoder-decoder-discriminator* as is represented in Figure 2.a and Figure 2.b. This architecture was designed so that the output image is coherent with the input image. The encoder reduces the dimensionality of the image to capture important features in a more compressed format, while the decoder undertakes the task of reconstructing the image from the aforementioned features, ensuring that the important attributes of the image are preserved with the best possible accuracy. The parameter  $\lambda$  is a weighting parameter that balances the importance of the L1 loss and the adversarial loss in the objective function of the Pix2Pix model, in other words, this value effectively balances the two components and reduces visual artifacts in various applications. Following the recommendation of the model's paper,  $\lambda$  was set to 100, as this value effectively balances the two loss components and reduces visual artifacts [23]. In comparison, the discriminator, a crucial component of the architecture, evaluates whether the generated images are distinguishable from real images, thereby providing feedback during the training process. An important feature of this architecture is the use of skip connections, which, as the name suggests, skip one or more layers in the cGAN and feed the output of one layer to the layer following the skipped layer(s). In such a way that it can be expressed under the following configuration:

**encoder:**

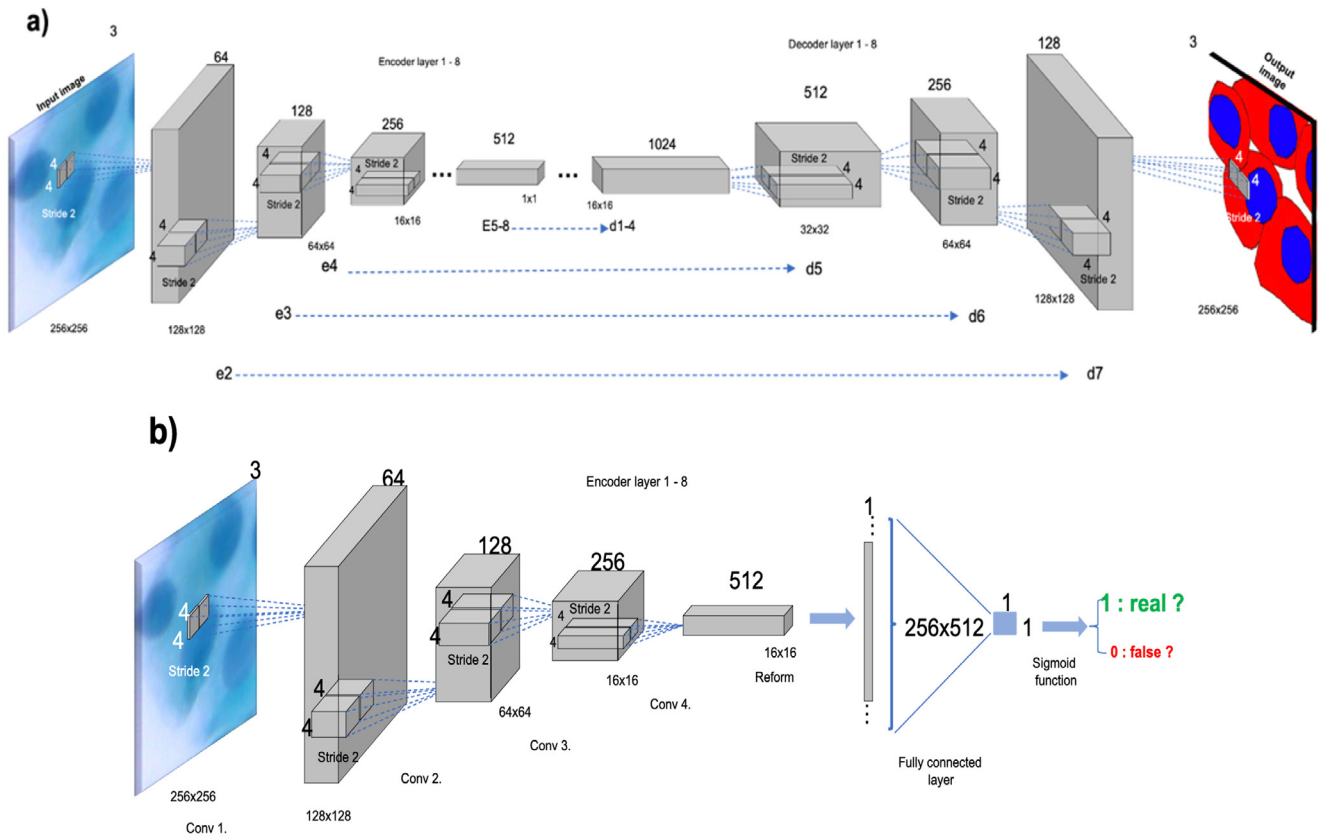
C64-C128-C256-C512-C512-C512-C512

**decoder:**

CD512-CD512-CD512-C512-C256-C128-C64

**discriminator:**

C64-C128-C256-C512-C512-C512



**FIGURE 2. Generator and discriminator architecture for segmentation of cervical intraepithelial nuclei. (a) Generator architecture. Encoder, decoder and the skip connections. (b) Discriminator Architecture.**

Where  $Ck$  is a Convolution-BatchNorm-ReLU layer with  $k$  filters.  $CDk$  refers to a Convolution-BatchNorm-Dropout-ReLU layer, recommending a 50 % dropout rate and 6x6 spatial filters, with a stride of 2. The model recommends 4x4 filters, however this parameter was changed to 6x6 in order to detect more extensive spatial features. Regarding the decoder in its last layer, a convolution layer must be made to match the pathways of color, in this case there are 3. Finally, a layer with an activation function of Hyperbolic tangent was added. The BatchNorm application was applied as an exception in the first C64 layer of the encoder.

### Evaluation metrics

An approach based on pixel comparison was used to evaluate the images segmented by the Pix2Pix model and the real segmented images. These metrics provide a holistic understanding of model performance to segment both nuclei and cytoplasm.

**TABLE 2. Multiple confusion matrix.**

	Predicted positive	Predicted negative
Actual Positive	TP	FN
Actual Negative	FP	TN

Where **TP** (True Positive) are the correctly segmented pixels as nuclei or cytoplasm, **FP** (False Positive) are the pixels erroneously segmented as nuclei or cytoplasm, **FN** (False Negative) are the pixels that were not segmented but should have been, and **TN** (True Negative) are the pixels that represent the pixels correctly identified as not belonging to the region of interest, this is best shown in Table 2.

From the data in the multiple confusion table, 5 important parameters can be calculated <sup>[26][27]</sup>.

**Sensitivity:** this measure indicates the percentage of the cases that the model identify as positive actually are, as defined in Equation (6).

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

**Specificity:** measures the model's ability to correctly identify negative cases. It is complementary to specificity and is essential in situations where correct identification of negatives is critical, as presented in Equation (7).

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

**Accuracy:** it measures the overall correctness of the model. It is the proportion of correctly classified instances (TP and TN) out of the total number of instances, as can be seen in Equation (8).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (8)$$

**Precision:** Accuracy focuses on how many of the predicted positive results were actually correctly identified. It is critical when false positives need to be minimized, as shown in Equation (9).

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

**Jaccard index (IoU):** is a spatial metric used in image segmentation. It calculates the overlap between predicted and actual areas of interest, as observed in Equation (10).

$$IoU = \frac{TP}{TP + FP + FN} \quad (10)$$

Finally, statistical measures such as mean and standard deviation are calculated for the results obtained from the evaluation metrics during the experiments. These measures allow for a more precise visualization of both the variability and consistency of the model's performance, thereby providing a robust measure of its effectiveness and capability.

## Experimentation and results

The experimental workflow was structured into several key stages to ensure accurate segmentation of nuclei and cytoplasm, as well as their evaluation in Pap smear images, using the Pix2Pix neural network. The experimentation in this study focused on segmenting and automatically evaluating nuclei and cytoplasm in the images.

### Data and implementation details

The dataset used contains information on the position of cell edges for each category, which were manually segmented by a professional. This manual process is crucial as it allows for the automatic generation of target images that not only accurately reflect the morphological characteristics of the cells from each category, but also ensure the representativeness and authenticity of the target image. The manual segmentation performed by an expert provided accurate boundary information.

From the original images, with a resolution of 2048x1536, the cores were cropped into images with a size of 256x256. With the same proportion of the coordinated segmentation dataset, images were created with the segmented cells in which the nucleus was assigned a color in the RGB format: blue (R:0, G:0, B:255), the cytoplasm in red (R:255, G:0, B:0) and finally the background white (R:255, G:255, B:255). In the same way, both the cytoplasm and the nucleus were joined by a black border (R:0, G:0, B:0). This can be seen in detail in Figure 3. Cells that were close to the edge were discarded, as information would be lost and/or could not be adjusted to the required size of 256x256. Of the total, 4049, single cell images available, only 3989 were used, with the Superficial/Intermediate, Parabasal, Koilocytotic, Metaplastic and Dyskeratotic categories.

The dataset was divided into two categories; 70 % of the images cropped from each of the categories of the dataset were taken for training. These images were taken randomly, in the same way the remaining 30 % were assigned for testing the Pix2Pix model, leaving a total of 2,792 for training and 1,197 for testing.

### Model training

Having the architecture as described in the previous section, it proceeded to train the model with a total of 200 *epochs* and the test images of the SIPaKMeD dataset. This number of epochs was selected based on recommendations observed in state-of-the-art studies, where it has been demonstrated to achieve optimal balance between training convergence and overfitting<sup>[23]</sup>. The Pix2Pix model needs both a reference image  $x$  and a target image  $y$ . These references are used to learn the relationship between the corresponding input and output image in a training dataset. That is why a target dataset  $Z$  was created after the original dataset, which represents the segmentation of the nuclei of each of the cells, as illustrated in Figure 3.

Data augmentation was carried out by using techniques that allow obtaining a greater representation of the information (images). Arbitrary changes of image rotation, horizontal and vertical flipping, zoom-in and zoom-out were implemented for variations in scale and random cuts in the image to show different parts. This technique, known as Jitter, allows the algorithm to have diverse information from a single image.

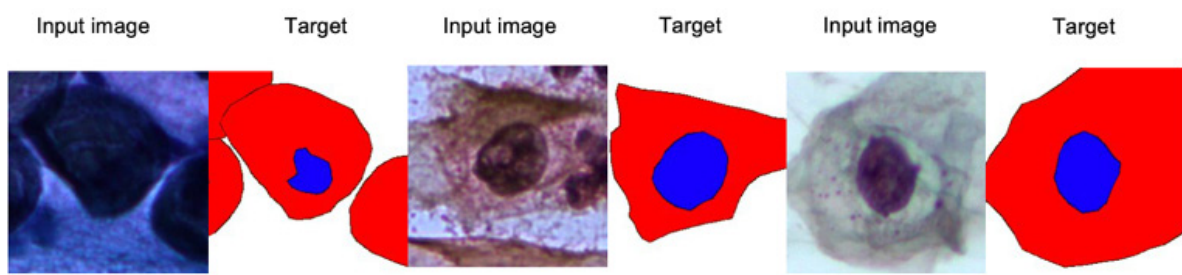


FIGURE 3. Image Samples and their output target

### Evaluation for generated images

Once the model training was completed, it was used to generate segmentation maps for the validation set, which corresponds to 30 % of the images. Regarding the evaluation of the model, quantitative and qualitative evaluation was carried out; the first consists of superimposing the segmented images onto the images produced by the model, counting the number of pixels of the color assigned to the nucleus, cytoplasm and background; ignoring the components, the three-color pathways, that are foreign to the object. In the same way, a threshold of 200 was applied on the colors corresponding to the classifications because the algorithm does not always assign the exact corresponding color to the pixel, varying slightly. This threshold was chosen after visual inspection of the images generated by the model and based on empirical observations, as it provided the best balance between including relevant pixels within the area of interest and excluding background noise or irrelevant variations, e.g. material not belonging to the area of interest. Furthermore, this threshold ensures that subtle variations in pixel intensity, inherent to the generative process, do not compromise the integrity of the segmentation process. This allows to have the most information about the area of interest, discarding areas with non-uniform colors or with a low percentage of target color.

The color images generated by the model were converted to a binary domain, both the nucleus and the cytoplasm part, in order to perform a logical AND operation between all the pixels and making a sum of the number of pixels that were correctly segmented with respect to their class. As represented in Figure 4, the pixels outside the object's area were calculated in the same way, this will give the percentage of segmentation error of the nucleus and cytoplasm and its false positives.

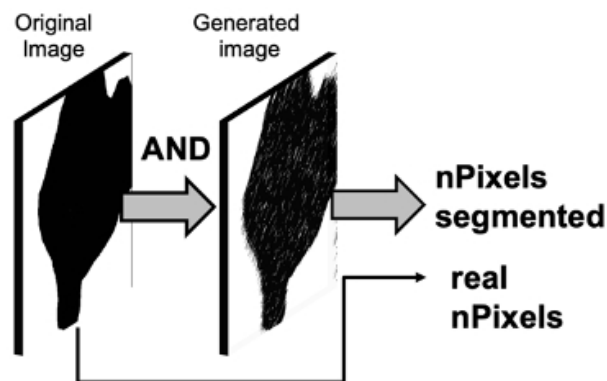


FIGURE 4. Quantitative evaluation of the generated image.

The metrics were calculated for each of the generated images, obtaining a set of data that measures central tendency as mean. As for the qualitative evaluation, the quality of the generated image is carried out. This may give indications if it is necessary to implement other algorithms in the future so that the segmented objects have a quality that allows obtaining morphological data and texture statistical data, among others.

## MATERIALS AND METHODS

In this study, the Pix2Pix model is evaluated for the segmentation of nuclei and cytoplasm of cells in Papanicolaou test images. Evaluation metrics such as accuracy, precision, sensitivity, specificity, and Jaccard index were used to assess the model. The results of these metrics for both nucleus and cytoplasm are shown in more detail in Table 3.

Regarding the segmentation of nuclei, the model achieved an accuracy of 93.99 %, a precision of 88.80 %, and a sensitivity of 89.62 %. These values demonstrate optimal performance in identifying and segmenting this area. Concerning the ability to distinguish between the background and the nucleus, the specificity reached 91.32 %, which is significant as it greatly reduces the false positive rate. The Jaccard index was 80.99 %, which quantifies the overlap between the predicted and true nucleus, indicating a satisfactory degree of overlap between the two areas.

On the other hand, for the cytoplasmic area, the Pix2Pix model achieved an accuracy of 96.93 %, a precision of 97.33 %, and a high sensitivity of 99.34 %. These results suggest that the model is highly capable of detecting cytoplasmic areas. However, the model performed significantly lower in terms of specificity compared to the other indicators, with a value of 67.03 %, this implies a significant rate of false positives. Although specificity was low, the Jaccard index reached 96.71 %, indicating substantial overlap between the predicted and true cytoplasmic areas.

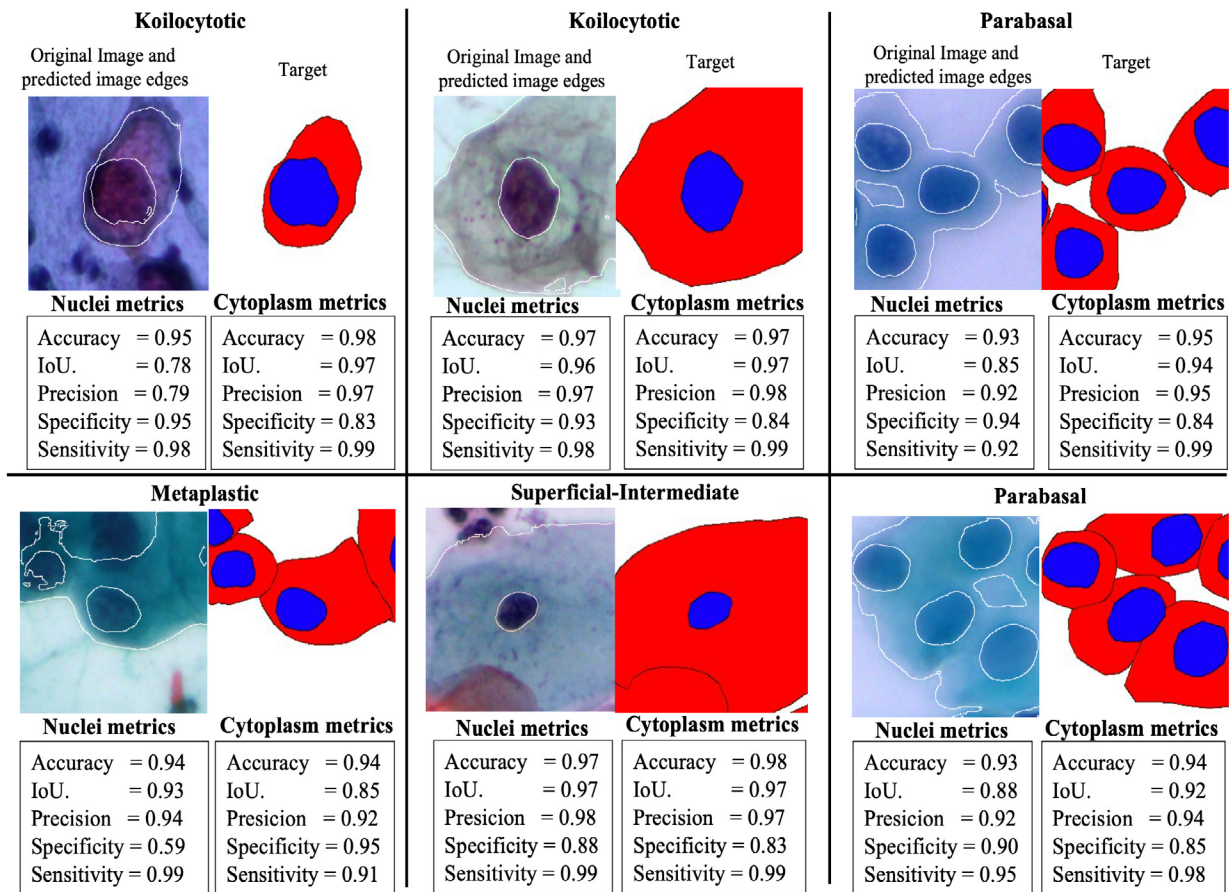
Visual and quantitative comparisons can be observed in Figure 5. The edges of the images predicted by the model are on the original image overlapping, next to its target and below these figures, the evaluation metrics for each example. It is important to note visually that the edges coincide in a significant percentage with the original images, which is of utmost importance to avoid false positives when performing evaluations to determine the level of neoplasia in the samples.

TABLE 3. Mean results for evaluation metrics.

Metric	Nuclei mean	Cytoplasm mean
Accuracy	93.99 %	96.93 %
Precision	88.80 %	97.33 %
Sensitivity	89.62 %	99.34 %
Specificity	91.32 %	67.03 %
Jaccard Index IoU)	80.89 %	96.71 %

TABLE 4. Standard deviation results for evaluation metrics.

Metric	Std. Dev. Nucleus	Std. Dev. Cytoplasm
Accuracy	5.12 %	3.17 %
Precision	10.68 %	3.20 %
Sensitivity	13.22 %	0.80 %
Specificity	11.64 %	12.01 %
Jaccard Index IoU)	15.89 %	3.33 %



**FIGURE 5.** Comparison with test image, in left-to-right order: the input image with predicted image edge and target. The evaluation metrics are shown below each image and the category above the figures.

The results obtained and shown in Table 3 indicate that, although the model performs well in detecting and segmenting areas of interest, there is a noticeable difference in specificity in the cytoplasmic area. This may be due to the complexity of these areas, taking into account whether the sample was carried out correctly, the overlapping of cells, or the level of neoplasia, as in certain stages of CIN, the cells tend to group and overlap. In Table 4 is shown the standard deviations of evaluation metrics, the model demonstrates more consistent segmentation performance for the cytoplasm across all metrics. In contrast, the higher variability in nucleus-related metrics highlights a need for further refinement, especially in sensitivity and Jaccard Index, to achieve more stable segmentation for this region. This could lead to further studies aimed at addressing this specific issue, incorporating preprocessing to highlight the cytoplasmic edges or training with more diverse data.

Table 5 compares the performance of different methods and models for the segmentation of nuclei and cytoplasm, using the evaluation metrics considered in this study. Pix2Pix stands out as one of the best-performing methods. The ConvNet model achieves the highest precision and specificity, with values of 0.98. On the other hand, U-Net and EfficientNet demonstrate excellent sensitivity, reaching 0.95 and 0.91, respectively. In contrast, Superpixel + FCN approaches show a specificity of 0.96 but fall short in sensitivity and Jaccard Index. Compared to other methods in the table, the precision and IoU of the Pix2Pix model yield optimal results. Lastly, the review <sup>[38]</sup> presents several methods for the segmentation and classification of Pap Smear images. While these models and algorithms are effective, the evaluation metrics used differ from those employed in this study.

TABLE 5. Segmentation techniques performance comparison.

Method	Nucleus and Cytoplasm				
	Accuracy	Precision	Sensitivity	Specificity	Jaccard Index IoU)
K-means [28]	-	-	-	-	0,72
Mean-shift	-	-	0,94	0,93	-
CNN [29]	-	0,73	0,65	0,73	-
Bi-Path[30]	0,77	-	0,72	0,71	-
<b>Pix2Pix</b> [23]	0,95	<b>0,93</b>	0,94	0,79	<b>0,89</b>
Superpixel and CNN [31]	-	-	0,87	0,91	
U-Net Model [32]	-	-	<b>0,95</b>	0,75	0,63
Otsu [33]	-	-	0,89	0,43	0,52
Star-Convex Polygons [34]	-	0,89	-	-	-
ConvNet [35]	<b>0,98</b>	-	-	<b>0,98</b>	-
Ensemble of(Superpixel, FCN8, FCN16) using STAPLE) [36]	0,87	-	0,66	0,96	0,53
EfficientNet combined with FPN [37]	-	0,91	0,91	0,95	0,84

## CONCLUSIONS

The high precision and sensitivity are a good indication that the Pix2Pix model is capable of segmenting both nuclei and cytoplasm. The standard deviation of Sensitivity and Accuracy were low with values of 0.1068 and 0.1322 respectively, this could indicate that the model has a homogeneous performance when identifying cell nuclei. Similarly, the low standard deviation values for Sensitivity and Accuracy in terms of cytoplasm were 0.032 and 0.008, respectively.

The model was trained with images containing the original resolution. The training with the original resolution took too long to complete, with a high computational cost. This resulted in low resolution output by the model, rendering the images to not correspond to the objective. In addition, it should be noted that the dataset used does not have the information of all the existing cells in the image. This generated a low adjustment, since incomplete information was entered. In other words, cells that should be segmented did not enter in that way.

The Pix2Pix model demonstrates balanced performance across all metrics, making it a solid choice for segmenting nuclei and cytoplasm. While models such as ConvNet achieve slightly higher accuracy and specificity, Pix2Pix excels in sensitivity and Jaccard index, which are critical for accurate segmentation tasks. This positions Pix2Pix as an effective and reliable method for applications requiring accurate and consistent segmentation in biomedical images.

The output images will be used in future work, which will obtain image characteristics such as morphological, texture, color, among others. The analysis with advanced detailed machine learning techniques will lead to evaluating and identifying in a concise manner the different levels of CIN that the images could present. This analysis will represent a more precise support for the present work and will contribute to the research and development of more precise and sophisticated diagnostic tools. In this same way, other existing architectures in the literature could be explored and comparisons made between them.

This model could be implemented in cloud applications or systems for CAD in remote or hard-to-reach locations, in this way images taken by specialized technicians in this area could be automatically processed and data analysis performed. This data can be saved and accessed anywhere by medical specialists in order to offer an effective, prompt and accurate diagnosis.

Within the limitations of the SIPaKMeD dataset is that not all the cells present in the images are segmented as individual cell images, which means there is no information about their location. This can result in incomplete data during the training phase, especially if one or more cells are not identified by specialists. Therefore, it is crucial to consider this limitation in future applications of models such as Pix2Pix.

In summary, the Pix2Pix Conditional Adversarial Generative Neural Network model is a powerful tool for segmentation of cell nuclei and cytoplasm in Pap smear images. This model leverages the power of deep learning to distinguish complex cells with high accuracy, enabling detailed analysis of cellular anomalies.

## **ETHICAL STATEMENT**

In this research, the SIPaKMeD dataset was used, which is publicly available and intended for research purposes. It consists of anonymized images of cervical cells and does not contain any information that could identify the patients. The dataset was specifically designed for research use, so obtaining additional informed consent was not necessary. No further ethical considerations were applied, as the nature of the dataset complies with regulations on anonymized data in research.

## **AUTHOR CONTRIBUTIONS**

F. J. C.-C. conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing original draft, writing - review & editing; J. M. C.P. data curation, supervision; C. E. G.-T. investigation, methodology, supervision, writing original draft, writing - review & editing; E. A.-C investigation, supervision, and writing - review & editing. All authors reviewed and approved the final version of the manuscript before submission.

## REFERENCIAS

- [1] American Cancer Society. "What is cervical cancer?" Accessed: Nov. 05, 2023. [Online]. Available: <https://www.cancer.org/cancer/types/cervical-cancer/about/what-is-cervical-cancer.html>
- [2] World Health Organization. "Cervical cancer." Accessed: Jul. 10, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>
- [3] The American Cancer Society. "Understanding what cancer is: ancient times to present." Accessed: Jul. 11, 2024. [Online]. Available: <https://www.cancer.org/cancer/understanding-cancer/history-of-cancer/what-is-cancer.html>
- [4] National Cancer Institute. "Cervical cancer—health professional version." Accessed: Jul. 10, 2024. [Online]. Available: <https://www.cancer.gov/types/cervical/hp>
- [5] The American Cancer Society, "Test for cervical Cancer" Accessed: Jul. 11, 2024. [Online]. Available: <https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/screening-tests.html>
- [6] J. F. Barter, "The life and contributions of Doctor George Nicholas Papanicolaou," *Surg. Gynecol. Obstet.*, vol. 174, no. 6, pp. 530-532, 1992.
- [7] National Cancer Institute. "HPV and Pap test results: next steps after an abnormal cervical cancer screening test." Accessed: Jul. 19, 2024. [Online]. Available: <https://www.cancer.gov/types/cervical/screening/abnormal-hpv-pap-test-results>
- [8] J.W. Sellors and R. Sankaranarayanan, *Colposcopy and treatment of cervical intraepithelial neoplasia: a beginners manual*, Cervical Cancer Screening - Chapter 2. (2003). Accessed: Aug. 20, 2024. [Online]. Available: <https://screening.iarc.fr/colpochap.php?lang=3&chap=2>
- [9] R. Nayar and D. C. Wilbur, *The Bethesda system for reporting cervical cytology: definitions, criteria, and explanatory notes*, 3rd. Ed., Springer, 2015, doi: <https://doi.org/10.1007/978-3-319-11074-5>
- [10] R. A. Castellino, "Computer aided detection (CAD): an overview," *Cancer Imaging*, vol. 5, no. 1, pp. 17-19, 2005, doi: <https://doi.org/10.1102/1470-7330.2005.0018>
- [11] P. Wang et al., "Automatic cell nuclei segmentation and classification of cervical Pap smear images," *Biomed. Signal Process Control*, vol. 48, pp. 93-103, 2019, doi: <https://doi.org/10.1016/j.bspc.2018.09.008>
- [12] M. Zhao et al., "SEENS: Nuclei segmentation in Pap smear images with selective edge enhancement," *Future Gener. Comput. Syst.*, vol. 114, pp. 185-194, 2021, doi: <https://doi.org/10.1016/j.future.2020.07.045>
- [13] R. Saha, M. Bajger, and G. Lee, "Spatial shape constrained Fuzzy C-Means (FCM) clustering for nucleus segmentation in pap smear images," in 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia, 2016, pp. 1-8. doi: <https://doi.org/10.1109/DICTA.2016.7797086>
- [14] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognit.*, vol. 34, no. 12, pp. 2259-2281, 2001, doi: [https://doi.org/10.1016/S0031-3203\(00\)00149-7](https://doi.org/10.1016/S0031-3203(00)00149-7)
- [15] M. Lucic et al., "Are GANs created equal? a large-scale study," 2017, arXiv: 1711.10337, doi: <https://doi.org/10.48550/arXiv.1711.10337>
- [16] S. Kato and K. Hotta, "Cell segmentation by image-to-image translation using multiple different discriminators," in Proc. 13th Int. Joint Conf. Biomed. Eng. Sys. Technol., SCITEPRESS -Science and Technology Publications, Valletta, Malta, 2020, vol. 4, pp. 330-335, doi: <https://doi.org/10.5220/0009170103300335>
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in 18th International Conference Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, in Lecture Notes in Computer Science (LNIP), vol. 9351, 2015, pp. 234-241, doi: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [18] J. Zak et al., "Cell image augmentation for classification task using GANs on Pap smear dataset," *Biocybern. Biomed. Eng.*, vol. 42, no. 3, pp. 995-1011, 2022, doi: <https://doi.org/10.1016/j.bbe.2022.07.003>
- [19] S. Tahri et al., "A high-performance method of deep learning for prostate MR-only radiotherapy planning using an optimized Pix2Pix architecture," *Phys. Med.*, vol. 103, pp. 108-118, 2022, doi: <https://doi.org/10.1016/j.ejmp.2022.10.003>
- [20] Y. Xue et al., "Selective synthetic augmentation with HistoGAN for improved histopathology image classification," *Med. Image Anal.*, vol. 67, 2021, art. no. 101816, doi: <https://doi.org/10.1016/j.media.2020.101816>
- [21] W. A. Mustafa, A. Halim, M. A. Jamlos, and S. Z. S. Idrus, "A Review: Pap Smear Analysis Based on Image Processing Approach," in 2nd Joint International Conference on Emerging Computing Technology and Sports (JICETS), Bandung, Indonesia, 2019, vol. 1529, art. no. 22080, doi: <https://doi.org/10.1088/1742-6596/1529/2/022080>
- [22] M. E. Plissiti et al., "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), Athens, Greece, 2018, pp. 3144-3148, doi: <https://doi.org/10.1109/ICIP.2018.8451588>
- [23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2016, arXiv: 1611.07004, doi: <https://doi.org/10.48550/arXiv.1611.07004>
- [24] I. J. Goodfellow et al., "Generative Adversarial Networks," 2014, arXiv: 1406.2661, doi: <https://doi.org/10.48550/arXiv.1406.2661>
- [25] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," 2014, arXiv: 1411.1784, doi: <https://doi.org/10.48550/arXiv.1411.1784>
- [26] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, arXiv: 2010.16061, doi: <https://doi.org/10.48550/arXiv.2010.16061>

- [27] M. Sokolova and G. Lalpalmé, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427-437, 2009, doi: <https://doi.org/10.1016/j.ipm.2009.03.002>
- [28] M. Neghina et al., "Automatic detection of cervical cells in Pap-smear images using polar transform and k-means segmentation," in 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 2016, pp. 1-6, doi: <https://doi.org/10.1109/IPTA.2016.7821038>
- [29] Y. Song et al., "Accurate Cervical Cell Segmentation from Overlapping Clumps in Pap Smear Images," *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 288-300, 2017, doi: <https://doi.org/10.1109/TMI.2016.2606380>
- [30] A. Desiani et al., "Bi-path architecture of CNN segmentation and classification method for cervical cancer disorders based on Pap-smear images.," *IAENG Int. J. Comput. Sci.*, vol. 48, no. 3, 2021. [Online]. Available: [https://www.iaeng.org/IJCS/issues\\_v48/issue\\_3/IJCS\\_48\\_3\\_37.pdf](https://www.iaeng.org/IJCS/issues_v48/issue_3/IJCS_48_3_37.pdf)
- [31] Y. Song et al., "A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 2014, pp. 2903-2906, doi: <https://doi.org/10.1109/EMBC.2014.6944230>
- [32] N. Nazir, A. Sarwar, B. S. Saini, and R. Shams, "A Robust deep learning approach for accurate segmentation of cytoplasm and nucleus in noisy pap smear images," *Computation*, vol. 11, no. 10, 2023, art. no. 195, doi: <https://doi.org/10.3390/computation11100195>
- [33] C. Huang, X. Li, and Y. Wen, "AN OTSU image segmentation based on fruitfly optimization algorithm," *Alex. Eng. J.*, vol. 60, no. 1, pp. 183-188, 2021, doi: <https://doi.org/10.1016/j.aej.2020.06.054>
- [34] Y. Zhao et al., "Automatic segmentation of cervical cells based on star-convex polygons in Pap smear images," *Bioengineering*, vol. 10, no. 1, 2022, art. no. 47, doi: <https://doi.org/10.3390/bioengineering10010047>
- [35] L. Zhang et al., "DeepPap: deep convolutional networks for cervical cell classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1633-1643, 2017, doi: <https://doi.org/10.1109/JBHI.2017.2705583>
- [36] B. Harangi et al., "Cell detection on digitized Pap smear images using ensemble of conventional image processing and deep learning techniques," in 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 2019, pp. 38-42, doi: <https://doi.org/10.1109/ISPA.2019.8868683>
- [37] M. J. Del Moral-Argumedo, C. A. Ochoa-Zezzati, R. Posada-Gómez, and A. A. Aguilar-Lasserre, "A deep learning approach for automated cytoplasm and nuclei cervical segmentation," *Biomed. Signal. Process. Control.*, vol. 81, 2023, art. no. 104483, doi: <https://doi.org/10.1016/j.bspc.2022.104483>
- [38] B. Z. Wubineh, A. Rusiecki, and K. Halawa, "Segmentation and classification techniques for Pap smear images in detecting cervical cancer: a systematic review," *IEEE Access*, vol. 12, pp. 118195-118213, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3447887>