UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica
https://doi.org/10.14483/issn.2248-4728

VISIÓN ELECTRÓNICA

A RESEARCH VISION

Development of machine learning models for binary quality classification and noise removal in Pap smear samples

Desarrollo de modelos de machine learning para clasificación binaria de calidad y eliminación de ruido en muestras de Papanicolau

Heidy Vanessa Nuñez-Tovar[1], Diana Lorena Forero Guevara[2]

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:
Enviado: 01/07/2025
Recibido: 01/07/2025
Aceptado: 17/11/2025

Keywords:
Cervical cytology
Diffusion model
Diffusion model

Palabras clave:
Citología cervical
Modelo de difusión
Transfer learning

ABSTRACT

The subjectivity and agility in the review and quality classification of cervical cytology images represents a significant challenge due to the individual observer's criteria, as well as the high volume of samples requiring analysis. The project aims to develop two machine learning models; the first is a classification model that categorizes digitized samples as satisfactory or unsatisfactory. The MobileNet, VGG16, and Resnet50 architectures were compared, yielding better results with the latter, reaching a sensitivity of 0.93 for unsatisfactory samples. The second, a diffusion model for noise reduction where a UNet architecture with ResNet blocks was evaluated for images without noise and with added noise, and an unsharp mask was applied, achieving PSNR and SSIM metrics of 36 dB and 0.92 in noise-free images, and 31 dB and 0.72 in noisy images. The implementation of these models serves as a first step in the binary classification of cytological image quality and in improving the initial image quality.

RESUMEN

La subjetividad y la agilidad en la revisión y clasificación de calidad de imágenes de citología cervical representa un desafío importante debido al criterio individual del observador, así como al alto volumen de muestras que requieren análisis. El proyecto tiene como objetivo desarrollar dos modelos de machine learning; el primero es un modelo de clasificación que categoriza las muestras digitalizadas como satisfactorias o insatisfactorias, se compararon las arquitecturas MobileNet, VGG16 y Resnet50, arrojando mejores resultados con esta última, llegando a una sensibilidad de 0.93 las muestras insatisfactorias. El segundo, un modelo de difusión para reducción de ruido donde se utilizó una arquitectura UNet con bloques ResNet...

1. Licenciada en Matemáticas de la Universidad Surcolombiana, Colombia. Especialista en Estadística de la Universidad Surcolombiana, Colombia. Magister en Estadística Aplicada y Ciencias de Datos de la Universidad El Bosque, Colombia. E-mail: hnunez@unbosque.edu.co. ORCID: https://orcid.org/0000-0001-8308-7815
2. Ingeniera Industrial de Universidad Militar Nueva Granada, Colombia. Magister en Estadística Aplicada y Ciencia de Datos de Universidad El Bosque, Colombia. E-mail: dlforero@unbosque,edu.co ORCID: https://orcid.org/0000-0003-4507-2035

Cite this article as: A. F. Jiménez-López, F. R. Jiménez-López, C. S. León-Álvarez, "Surveillance of insect pests in corn crops: monitoring and counting using electronic traps", Visión Electrónica, vol. 19, no. 1, 2025.

1. Introduction

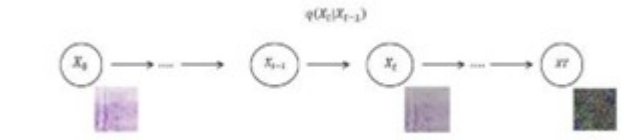
Cervical cancer is a common sexually transmitted disease caused by infection with the human papillomavirus (HPV) [1]. To world levelIt is the eighth cancer with the highest incidence, the ninth with the highest mortality and a prevalence in the last 5 years of approximately 50 cases per 100,000 people [2]; in Colombia it is the third most common cancer in women [3] One of the prevention methods is performing the Pap test for cervical cytology, which is the main screening test for cervical cancer, contributing to early detection. Review of Pap samples is performed manually using an optical microscope under the level of expertise and subjectivity of the examining pathologist, which can lead to false positives or false negatives, thus affecting the patient's diagnosis and treatment [4]. The impact of Artificial Intelligence (AI) in the health sector has been significant in the early detection of diseases and reducing the burden on professionals, through the implementation of machine learning models, for example, in [5] and in [6] They develop algorithms with neural networks to detect cervical cancer, demonstrating the capacity of AI in optimizing medical diagnoses. Training these models involves a large volume and variety of scanned whole slide images (WSI), which are susceptible to factors that can affect their quality from the capture stage to the digitization of the sample. The latter can be influenced by factors such as different types of scanners, out-of-focus areas, inappropriate color profile, and unrecognized slide identifier [7], which can affect the model's performance. Therefore, algorithms that operate without prior quality analysis could generate inconsistent results [8], as in [9] where it is obtained a larger area under the curve in the models that used high-quality images. In recent studies [10] They have made progress in classifying the quality of cytological images for the detection of cervical cancer, with a categorization of poor and good quality; in another study [6], developed a quality system for images of thin-layer liquid-based cervical cell smears, categorized as satisfactory and unsatisfactory. Other works [11], have improved images in cervical cytology, using methodologies such as histogram equalization to adjust contrast, noise filtering and edge detection. This highlights the need for WSI image quality control as part of the diagnostic process, enabling detection and improvement of quality. Therefore, in this research, two machine learning models were developed to classify images as satisfactory and unsatisfactory. Noise was subsequently removed only for those images previously classified as satisfactory. Finally, a sharpening mask was applied. This contributes to a broader project focused on the early detection of cervical cancer in partnership with the Universidad El Bosque and the Colombian League Against Cancer in Bogotá.

2. Diffusion Models

2.1. Mathematical basis

Diffusion models are generative models that operate under the principles of forward and reverse diffusion [12]. This process is based on the structure of Markov chains, where each step depends only on the previous step; in diffusion models, noise is added in the forward diffusion process, where each state t depends on t-1 , and in the reverse process, noise is reversed, where each step t-1 depends on t the state. Thus, this process is based on conditional distribution, assuming a Gaussian distribution [13]. The following figure shows the diffusion model process:

Figure 1. Diffusion model process. Taken and modified from



a. Direct Diffusion

a. Reverse Diffusion

The forward diffusion process takes the image and adds noise in a series of steps, taking into account the variance value, which defines the amount of noise added. Equations (1) and (2) present the single-step process and the complete.

$$q(x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

$$q(x_0) = \prod_{t=1}^{T} q(x_{t-1}) \tag{2}$$

Here represents the normal (Gaussian) distribution, the image at step , is generated around the mean and variance I (1). In (2) represents the joint probability of the noisy images, the notation indicates the multiplication of probabilities of each individual step.

$$p\theta(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{3}$$

$$p\theta(X_t \mid X_{t-1}) := N(X_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(X_t, t)) \tag{4}$$

$p_\theta(x_{t-1} \mid x_t)$ represents the probability distribution of obtaining the previous image state x_{t-1} from the current

noisy image x_t during the reverse diffusion process. $\mathcal{N}$ denotes a Gaussian normal distribution, where $\mu_\theta(x_t, t)$ is the mean predicted by the model at step t, and $\Sigma_\theta(x_t, t)$ represents the variance of the distribution. The joint probability of the reverse diffusion process is defined as:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t) \tag{5}$$

where $p(x_T)$ corresponds to the initial noise distribution and the product term represents the sequence of conditional probability distributions applied at each diffusion step.

2.2. Image quality assessment metrics

The mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were used as image quality assessment metrics. The goal was to find the lowest MSE value and the highest PSNR and SSIM values.

2.2.1. MSE

The MSE together with the PSNR have been traditional metrics used as efficiency criteria in image filtering processes [14]. MSE measures signal fidelity by comparing the difference between an original image and the same image filtered pixel by pixel. MSE is a metric that is sensitive to changes in squaring the differences. It is simple and inexpensive to calculate [15] although it does not measure the perception of the quality of the visual image [16]. The calculation of the MSE is given by:

$$[MSE]_{(x.y)} = 1/N \sum_{i=1}^{N} [(x_i - y_i)]^2 \quad (5) \tag{6}$$

N is the total number of pixels in the images, and N is the values of the i-th points of the two signals or images.$x_i y_i$[17].

2.2.2. PSNR:

It is based on the MSE [18], and refers to the ratio between the maximum possible power of a signal (original image) and the power of the noise (reconstructed image). It is expressed in decibels, where an appropriate range is between 30 dB and 40 dB [19]. When an image has the same dynamic range, PSNR does not provide additional information to that obtained with MSE [20]. The PSNR calculation is given by:

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{L^2}{\mathrm{MSE}} \right) \tag{6}$$

L is the dynamic range of allowed image pixel intensities [20]

2.2.3. SSIM:

It measures three factors of images: luminance, contrast and structure.[21] In the process of removing noise from an image, changes to the structure of the image can be generated that are perceptible to human vision called structural distortions, and the work of this metric is this human function [17]. The SSIM range is from 0 to 1, where the highest values closest to 1 are sought. The SSIM calculation is given by:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\underline{x}\underline{y}}{(\underline{x})^2 + (\underline{y})^2} \cdot \frac{2\sigma_x \sigma_y}{{\sigma_x}^2 + {\sigma_y}^2} \tag{7}$$

The first part is the structure component, where the numerator is the covariance between the image and , and the bottom part is the standard deviation. The second part is the luminance, which represents the average intensity values for each image, and the third is the contrast component.xy [17].

3. Methodology

The proposed methodology is shown in the following figure.

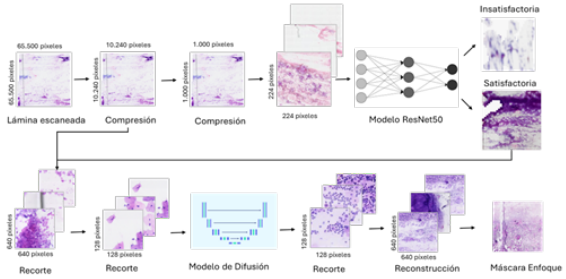


Figura 1: Methodology

3.1. Dataset and processing

For the visualization and cropping process of TIFF images, the TIAToolbox library in Python was used, which is focused on the analysis of pathological images **ref22**. Fifty cytology slides provided by the CITOMAP Cytology and Pathology Laboratory were selected. These samples were older than 5 years, so mounting the slides using synthetic resin (CYTORESIN) was required for proper cytology review by the cytologist and for the scanning process. The specialist reviewed the slides according to the 2014 Bethesda classification system, which is used for quality assessment and diagnosis. The samples were classified into two classes: satisfactory and unsatisfactory, taking into account factors such as fixation, staining, mounting, presence of endocervical cells, and presence of exocervical cells. The samples were then scanned by the University Foundation of Health Sciences (FUCS). It

should be noted that scan quality was included among the factors considered for the unsatisfactory class. This quality assessment was based on two variables: the number of white areas in the image and the percentage of blur. A sample was classified as unsatisfactory if it presented a hemorrhagic smear, inflammation, low cellularity, or poor scan quality. The samples were digitized using a MoticEasyScan scanner line and were initially stored in SVS format. However, as mentioned in **ref23**, this format is not compatible with some visualization software and its size is larger compared to other image formats. Therefore, TIFF format was selected for further processing. The scanned images had an original resolution of 65,500 $\times$ 65,500 pixels, a magnification of 40x, and a resolution of 5.3004 μm per pixel. To optimize storage and computational processing, the images were compressed using the Python PIL library, reducing their dimensions to 10,240 $\times$ 10,240 pixels, equivalent to approximately a 45 % reduction in file size. Initially, the images were compressed to 24,584 $\times$ 51,600 pixels using 1,024 processing blocks to avoid memory overflow problems. Subsequently, the final resolution of 10,240 $\times$ 10,240 pixels was obtained.

3.2. Quality classifier:

The VGG16, MobileNet, and ResNet50 architectures were evaluated, taking as reference the studies presented in **ref10** and **ref9**, where these architectures were implemented for medical image classification tasks. The referenced studies reported performance metrics of 99.5 % in sensitivity and precision for the first study, and an AUC of 99.9 % for the second study. The final layers of each architecture were customized to adapt the models to the binary classification task. Different layer configurations were tested, varying the number of trainable layers and the initial learning rate. Additionally, two callbacks were implemented during the training process. The first callback, `ModelCheckpoint`, was used to save the best model according to validation accuracy. The second callback, `ReduceLROnPlateau`, dynamically adjusted the learning rate during training. For the classifier model, 50 images with dimensions of 10,240 $\times$ 10,240 pixels were used. A preprocessing stage was performed in which the images were reduced to 1,000 $\times$ 1,000 pixels in JPG format. Subsequently, three crops of 224 $\times$ 224 pixels were extracted from each image to adapt them to the input size required by the evaluated architectures. This number of crops was selected to avoid ties between the two categories during model inference. A total of 150 image crops were obtained, distributed as 69 unsatisfactory samples and 81 satisfactory samples. For the training process, the dataset was divided into:

- 104 images for training
- 23 images for validation
- 23 images for testing

Data augmentation techniques were applied to all subsets, including rotations, translations, zooming, and horizontal flipping.

3.3. Probabilistic diffusion model for noise removal

For the diffusion model, images with dimensions of 10,240 x 10,240 pixels were taken, and cutouts with a size of 640 x 640 pixels were made in jpg format, generating 256 images per sheet, resulting in a total of 12,800 images for the final dataset. Taking into account the specifications of the diffusion model, 5 cutouts of 128 x 128 pixels were taken from each 640 x 640pixel image, subsequently in the final output of the model a spatial reconstruction was performed to return the images to their original size. The code that was implemented for the model was taken from the GitHub repository [24] which is available under the MIT license.

3.4. Unsharp Mask

Unsharp masking was implemented to improve image sharpness, making the images appear clearer and more defined.

$$\text{enhanced image} = \text{original} + \text{amount} * (\text{original} - \text{blurred}) \tag{8}$$

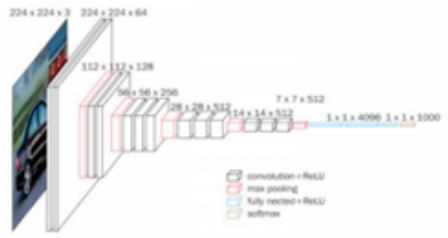
The process consists of taking the original image and calculating the difference between the original image and a blurred version of it. This difference is multiplied by the sharpening factor and then added back to the original image, resulting in a sharper image. For this purpose, the function `unsharp_mask` from the `skimage.filters` library was used **ref25**. This function receives two hyperparameters: `radius` and `amount`. The `radius` parameter determines the level of blur applied before edge enhancement, while the `amount` parameter controls the intensity of the sharpening effect.

4. Results

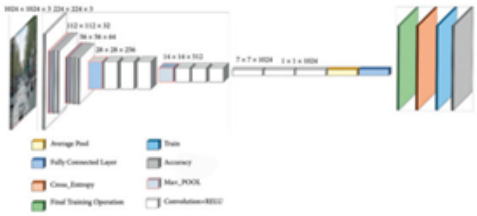
4.1. Classification model

For the selection of hyperparameters, the *Keras Tuner* methodology was implemented with the objective of obtaining the highest Recall value for the unsatisfactory class. This criterion was selected in order to reduce the number of false negatives, since an unsatisfactory sample incorrectly classified as satisfactory would be included in subsequent analyses

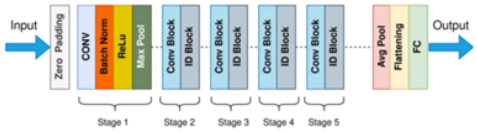
and incorporated into the dataset used to train the diagnostic model of the macroproject. Figure 2 presents the classical structure of each of the implemented architectures.



a. VGG16



b. MobileNetV1



c. ResNet50

Figura 2: Classical structure of the implemented architectures

In each of the related architectures, adjustments were made that allow the model to reduce overfitting and improve training efficiency. The classic VGG16 architecture contains 5 convolution blocks with 3×3 filters, increasing in number (64 to 512) per block and ReLU activations, each block has a MaxPooling layer, ending with 3 dense layers and a softmax layer. To adapt the training to the binary classifier, trainable layers were unfrozen, a `GlobalAveragePooling2D` layer was placed, an intermediate dense layer, a `Dropout` and the final layer adjusting to the output of 2 classes, as well as the optimizer learning rate. The classic MobileNetV1 architecture contains 28 depthwise separable convolutional layers with batch normalization and ReLU activations, a `GlobalAveragePooling` and `Dropout` layer, and a final `Dense` layer. Among the adjustments made for this architecture are the unfreezing of trainable layers, a `DropOut` layer, a `Flatten` layer, a final `Dense` layer, and a learning rate. The classic Resnet50 architecture starts with a 7×7 convolution layer and max-pooling, four stages of bottleneck residual blocks and finally a `GlobalAveragePooling` layer and a dense layer, as in the previous architectures, trainable layers were unfrozen, a `GlobalAveragePooling2D` layer was added, a `Dropout`, an intermediate dense layer and the final dense layer. The following tables list the results of each model trained with 90 epochs and the best hyperparameters, with the ResNet50 architecture performing best for the Recall of the unsatisfactory class with a value of 0.93.

Architecture	Configuration	Thawed layers	Learning rate

ResNet50	`GlobalAveragePooling2D`
	`Dropout(0.4)`
	`Dense(64, activation='relu')`
	`Dense(2, activation='softmax')`
		5	Adam 0.0000494

MobileNet	`Dropout(0.4)`
	`Flatten`
	`Dense(2, activation='softmax')`
		15	Adam 0.00029

VGG16	`GlobalAveragePooling2D`
	`Dropout(0.6)`
	`Dense(256, activation='relu')`
	`Dense(2, activation='softmax')`
		15	Adam 0.0000394

Cuadro 1: Better hyperparameters for architectures

		Precision	Recall	F1 score	# images
VGG16	*Class 1: unsatisfactory*	0.91	0.70	0.79	492
	Class 0: satisfactory	0.74	0.92	0.83	451
MobileNetV1	*Class 1: unsatisfactory*	0.85	0.83	0.84	492
	Class 0: satisfactory	0.82	0.84	0.83	451
ResNet50	*Class 1: unsatisfactory*	0.75	0.93	0.83	492
	Class 0: satisfactory	0.90	0.65	0.76	451

Cuadro 2: Results

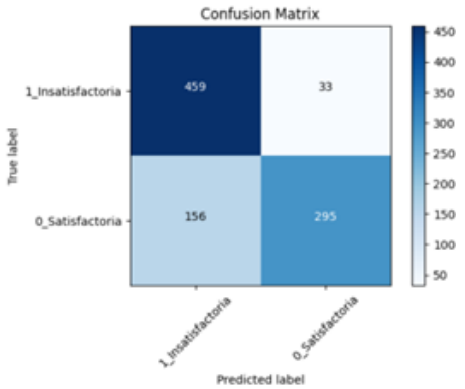


Figura 3: Confusion Matrix Best ResNet50 Model

4.2. Diffusion model

The model used was based on [24] where they take an adjusted HuggingFace model, which is based on a Classic UNet with advanced ResNet convolution blocks, time embeddings and linear attention. Table 3 shows the differences between Classic UNet and the implemented architecture:

Component	UNet Classic	HuggingFace
First Convolution	Conv 3x3	Conv 7x7
Block convolution	Conv + ReLU	ResNetBlock + SiLU
Time embeddings	Not applicable	If applicable
Linear attention	Not applicable	Linear care at each stage
Dimensionality reduction	Max Pooling 2x2	2D Stride 2 Convolution
Dimensionality increase	Up conv 2x2	Transposed convolution

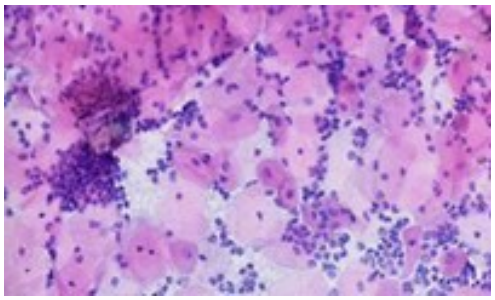
Cuadro 3: Differences between UNet Classic and HuggingFace Adjusted

The implemented architecture uses ResNetBlock convolutional blocks with a SiLU activation function and linear attention. These blocks consist of two 3x3 layers with normalization, which gives the model greater ability to detail the image; linear attention gives more weight to more important areas of the image without consuming too much memory. SiLU activation is more advanced and smoother compared to ReLU, allowing the model to learn complex patterns in a more controlled manner. In the first convolution, unlike classic UNet, the model used an input with a 7x7 kernel. In the downsampling phase, a 2D convolution layer with stride 2 is used to extract features while reducing the resolution. The bottleneck (the lowest part) uses ResNetBlock blocks; in the upsampling phase, a transposed convolution is used, doubling the resolution along with the advanced blocks. The entire architecture uses a temporal embedding function to add information about the time step in the diffusion process so that the model knows how much noise has been added to the image at that step. This allows the model to learn to progressively reconstruct the image by appropriately removing noise based on the passage of time. In the training process, different dataset size values were used in a range from 100 to 6,000, however, due to computational resource limitations, the best result was obtained with 2,307 images of 128x128 pixels which were divided into 1,600 images for the training set, 402 for the validation set and 350 for the test set. To create the diffusion datasets, a series of transformations were performed on the training set, consisting of converting the image to a tensor with values ranging from 0 to 1, randomly adjusting the image's brightness and hue, horizontally flipping the image, horizontally translating it, and scaling it. The objective of this was to enable the model to learn image characteristics independently of their variations. Gaussian noise level control was applied with variance ranging from 0 to 5e-3 in 1000 steps. For the test and test sets, the image was converted to a tensor and noise control was included. The model ran with 200 epochs, under a batch size of 32 for training and 64 for validation using the Google Colab Pro tool with an A100 GPU. The MSELoss loss function and an AdamW optimizer were used, the learning rate was adjusted at each step with a maximum value of 0.0005 and with a OneCycleLR cycle. For the model evaluation process, image reconstruction steps of 1, 5, 10, and 15 were used, and an average of the MSE, PSNR, and SSIM metrics was taken. In this process, the model was tested with the original 640x640 images without added noise, and images with added Gaussian noise with a mean of 0 and variance of 0.01 and salt and pepper noise with a percentage of 0.01. A total of 128 images were used for each of the tests. This was done in order to evaluate what level of noise the model could eliminate and the optimal step for the process. Below are the results of each test performed:

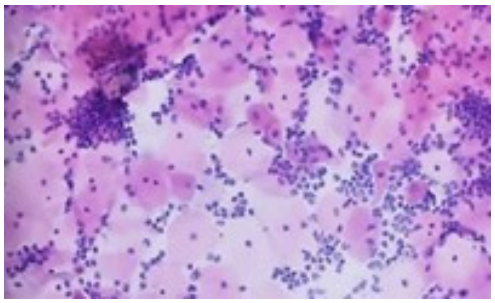
Image	Step	MSE	Manual PSNR (dB)	SSIM
Original image	1	27.61	34.99	0.90
Original image	5	25.20	35.52	0.91
Original image	10	22.06	36.11	0.92
Original image	15	15.57	37.29	0.94
Image with Gaussian noise	1	38.25	33.00	0.90
Image with Gaussian noise	5	37.83	32.89	0.90
Image with Gaussian noise	10	42.76	32.27	0.90
Image with Gaussian noise	15	48.97	31.55	0.83
Image with noise SaltPepper	1	43.12	31.87	0.62
Image with noise SaltPepper	5	42.90	31.88	0.60
Image with noise SaltPepper	10	43.58	31.79	0.56
Image with noise SaltPepper	15	47.16	31.41	0.49

Cuadro 4: Comparative metrics Diffusion Model

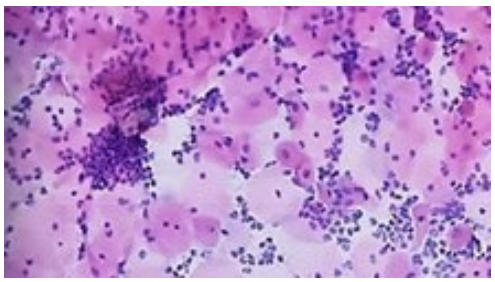
Given the obtained metrics, the discussions carried out with specialists, and the analysis of the processed images, it was concluded that step 15 was the optimal configuration since, although it removes a lower percentage of noise, it better preserves the characteristics of the images. To optimize image sharpness, an unsharp mask was applied at the selected step by varying the hyperparameters `radius` and `amount` using combinations of values of 1, 1.5, and 2. The combination of 1 and 1 provided the best visual perception for images without added noise, achieving MSE metrics of 21.9, PSNR values of 35.6 dB, and SSIM values of 0.92. For images with added Gaussian noise, the obtained metrics were 59.9 for MSE, 30.6 dB for PSNR, and 0.72 for SSIM. Regarding images with added salt-and-pepper noise, the unsharp mask test was not performed because the selected diffusion step did not demonstrate the best performance in noise removal. From this, it can be inferred that the model presented limitations because it was trained exclusively with Gaussian noise. Below is a fragment of an image processed by the diffusion model together with the unsharp mask.



a. Original image



b. MobileNetV1



c. ResNet50

Figura 4: Images without added noise processed by the Diffusion Model and Unsharp Mask

5. Conclusions

The two models deployed during the project were successfully developed, representing a significant advance in the selection and improvement of satisfactory images, contributing to the macro-project focused on cervical cancer detection. For the classifier model, the ResNet50 architecture proved to be the most suitable in terms of sensitivity to the unsatisfactory class compared to other architectures (VGG16 and MobileNet). The diffusion model based on the UNet architecture with ResNet blocks showed good results in noise removal, along with the application of an unsharp mask, which improved image sharpness. Two limitations were encountered during the project: first, the computational capacity available to run the diffusion model with a training size greater than 2,500 images. Second, since the diffusion model was trained with Gaussian noise, it was not highly effective in removing salt-and-pepper noise. Finally, it is suggested that higher zoom levels be used for future work with cytological images to increase practical applicability. Coauthors: Carlos

Puentes-Morales, Andres Felipe Mendoza-Cardona, Sandra Janneth Perdomo-Lara, Alex-Campos

Referencias

[1] World Health Organization, "Cervical cancer," Accessed: Nov. 21, 2024. [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/cervical-cancer

[2] International Agency for Research on Cancer, "World Fact Sheet," *Global Cancer Observatory*, 2022. [Online]. Available: https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf

[3] International Agency for Research on Cancer, "Colombia Fact Sheet," *Global Cancer Observatory*, 2022. [Online]. Available: https://gco.iarc.who.int/media/globocan/factsheets/populations/170-colombia-fact-sheet.pdf

[4] A. R. Bhatt, A. Ganatra, and K. Kotecha, "Cervical cancer detection in pap smear whole slide images using convNet with transfer learning and progressive resizing," *PeerJ Computer Science*, vol. 7, pp. 1–18, 2021, doi: https://doi.org/10.7717/peerj-cs.348.

[5] R. Gupta, A. Sarwar, and V. Sharma, "Screening of Cervical Cancer by Artificial Intelligence based Analysis of Digitized Papanicolaou-Smear Images," 2017. [Online]. Available: http://www.ijcmr.com

[6] X. Zhu et al., "Hybrid AI-assistive diagnostic model enables rapid TBS classification of cervical liquid-based thin-layer cell smears," *Nature Communications*, vol. 12, no. 1, 2021, doi: https://doi.org/10.1038/s41467-021-23913-3.

[7] R. Brixtel et al., "Whole Slide Image Quality in Digital Pathology: Review and Perspectives," *IEEE Access*, vol. 10, pp. 131005–131035, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3227437.

[8] R. J. Chalakkal, W. H. Abdulla, and S. S. Thulaseedharan, "Quality and content analysis of fundus images using deep learning," *Computers in Biology and Medicine*, vol. 108, pp. 317–331, May 2019, doi: https://doi.org/10.1016/j.compbiomed.2019.03.019.

[9] J. Wang et al., "Deep learning for quality assessment of retinal OCT images," *Biomedical Optics Express*, vol. 10, no. 12, p. 6057, Dec. 2019, doi: https://doi.org/10.1364/BOE.10.006057.

[10] T. Albuquerque et al., "Image Quality Assessment of Cytology Images using Deep Learning." [Online]. Available: https://www.researchgate.net/publication/345626844

[11] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images," *Biomedical Engineering Online*, vol. 18, no. 1, Feb. 2019, doi: https://doi.org/10.1186/s12938-019-0634-5.

[12] F. A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion Models in Vision: A Survey," Sep. 2022, doi: https://doi.org/10.1109/TPAMI.2023.3261988.

[13] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Jun. 2020. [Online]. Available: http://arxiv.org/abs/2006.11239

[14] E. Ieremeiev, V. Lukin, K. Okarma, and K. Egiazarian, "Full-reference quality metric based on neural network to assess the visual quality of remote sensing images," *Remote Sensing*, vol. 12, no. 15, Aug. 2020, doi: https://doi.org/10.3390/RS12152349.

[15] V. V. Lukin et al., "Weighted MSE Based Metrics for Characterization of Visual Quality of Image Denoising Methods," 2010. [Online]. Available: https://www.researchgate.net/publication/257259310

[16] G. Palubinskas, "Mystery Behind Similarity Measures MSE and SSIM," 2014.

[17] N. E. Mastorakis and V. Mladenov, *Advances in Visualization, Imaging and Simulation*. WSEAS Press, 2010.

[18] D. S. Turaga, Y. Chen, and J. Caviedes, "No reference PSNR estimation for compressed images," in *Signal Processing: Image Communication*, pp. 173–184, Feb. 2004, doi: https://doi.org/10.1016/j.image.2003.09.001.

[19] "Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation."

[20] P. Ndajah et al., "An investigation on the quality of denoised images," 2011. [Online]. Available: https://www.researchgate.net/publication/236897635

[21] D. R. I. M. Setiadi, "PSNR vs SSIM: imperceptibility quality assessment for image steganography," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8423–8444, Mar. 2021, doi: https://doi.org/10.1007/s11042-020-10035-z.

[22] J. Pocock et al., "TIAToolbox as an end-to-end library for advanced tissue image analytics," *Communications Medicine*, vol. 2, no. 1, 2022, doi: https://doi.org/10.1038/s43856-022-00186-5.

[23] "The SVS format — reaConverter." Accessed: Nov. 03, 2024. [Online]. Available: https://www.reaconverter.es/convert/svs.html

[24] "denoising-diffusion-models/denoising_diffusion_models.ipynb at main · EnricoPittini/denoising-diffusion-models · GitHub." Accessed: Nov. 03, 2024. [Online]. Available: https://github.com/EnricoPittini/denoising-diffusion-models/blob/main/denoising_diffusion_models.ipynb

[25] "Unsharp masking — skimage 0.24.0 documentation." Accessed: Nov. 03, 2024. [Online]. Available: https://scikit-image.org/docs/stable/auto_examples/filters/plot_unsharp_mask.html

[26] "Exploring ResNet50: An In-Depth Look at the Model Architecture and Code Implementation," Accessed: Nov. 03, 2024. [Online]. Available: https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-r

[27] K. D. Kadam, S. Ahirrao, and K. Kotecha, "Efficient Approach towards Detection and Identification of Copy Move and Image Splicing Forgeries Using Mask R-CNN with MobileNet V1," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi: https://doi.org/10.1155/2022/6845326.

[28] K. Kamal and H. EZ-ZAHRAOUY, "A comparison between the VGG16, VGG19 and ResNet50 architecture frameworks for classification of normal and CLAHE processed medical images," Apr. 2023, doi: https://doi.org/10.21203/rs.3.rs-2863523/v1.