Deep Learning and Transfer Learning Methods to Effectively Diagnose Cervical Cancer from Liquid-Based Cytology Pap Smear Images

https://doi.org/10.3991/ijoe.v19i04.37437

Lenis Wong[1(✉)], Andrés Ccopa[1], Elmer Diaz[1], Sergio Valcarcel[1], David Mauricio[1], Vladimir Villoslada[2]
[1]Professional School of Software Engineering, Faculty of System Engineering, Universidad Nacional Mayor de San Marcos, Lima, Perú
[2]Medicine Program, Faculty of Human Medicine, Universidad Nacional de Cajamarca, Cajamarca, Perú
lwongp@unmsm.edu.pe

Abstract—As cervical cancer is considered one of the leading causes of death for women globally, different screening techniques have emerged. As the Papanicolaou technique generates high numbers of false negatives due to only testing 20% of a sample, the liquid-based cytology technique was developed to test 100% of the sample and improve accuracy. However, as the larger sample size has made it difficult to detect the lesion images through a microscope, studies have looked for ways to intelligently analyze sample. The aim of this study is to develop an artificial intelligence image recognition system that detects the lesion level of cervical cancer of liquid-based Pap smears under the Bethesda classification of cancer (NILM/LSIL/HSIL/SCC). For this purpose, six activities were carried out: dataset selection, data augmentation, optimization, model development, evaluation and system construction. A dataset built from publicly available Pap smear images and passed through data augmentation algorithms generated a total of 2,676 images. Two models, ResNet50V2 and ResNet101V2, were developed under Deep Learning and Transfer Learning protocols. The evaluation showed that the ResNet50V2 model obtained better performance, where the classification of HSIL and SCC type images obtained a precision of 0.98 and achieved an accuracy of 0.97. Finally, the system based on the ResNet50V2 model was built and its performance was validated.

Keywords—cancer cervical, liquid-based cytology pap smear, deep learning, transfer learning, ResNet50V2, ResNet101V2

1 Introduction

Cervical cancer is a cellular alteration that originates in the epithelium of the cervix and initially manifests itself through precursor lesions that can slowly and progressivly evolve into cancer [1]. Studies during the period between 1974 and 1976 identified the mild, dysplastic lesions typical of Human Papilloma Virus (HPV) in the cervix as

a predominant risk factor for this cancer [2] and George Papanicolaou [3] proposed a method for early detection for their early detection. However, this method discards about 80% of the sample tissue, and high instances of false negatives occur when testing the reamaining 20%. This motivated researchers to search for new techniques, such as liquid-based cytology (LBC), which uses 100% of the sample [4]. Currently, cervical cancer is considered the fourth most common female malignancy globally and represents a major global health challenge [5].

Concurrently, artificial intelligence (AI) has emerged as a facilitator of what is being called smart medicine. As such, many studies apply AI image processing techniques for cervical cancer detection. Among them, [6] looked to classify three types of cervical cancer images by applying Transfer Learning (TL) methodologies. [7][8] proposed a convolutional neural network (CNN) architecture to classify images into two and seven classes. [9] applied image segmentation by building a SPFNet network to localize cancer cells. However, most of these studies are focused on conventional Pap smear images [6][7][8] or the localization of cancer cells [9]. In the present study, a working approach is proposed for the development of a decision support tool that uses Deep Learning (DL) and TL to analyze images generated by the liquid-based Papanicolaou tests in order to aide specialists, specifically cytologist pathologist, in the diagnosis of cervical cancer.

2 Background

2.1 Deep Learning (DL) and Transfer Learning (TL)

DL, also known as deep neural networks, is a subset of machine learning that involves multiple layers of data as opposed to just one. These DL networks do not need to be programmed with specific criteria in order to define an item, as they can define items autonomously simply by being exposed to large amounts of training data [10]. CNNs belong to a class of DL models that are prominently used in computer vision. TL is an approach commonly used for the creation of CNNs. It facilitates transferring knowledge acquired through problem solving to solve subsequent problems [6]. In this study, two CNN models are used by applying ImageNet weights [11][12] with an ImageNet dataset composed of more than 14 million images.

2.2 Convolutional neural network (CNN) and Residual Network (ResNet)

In any neural network, the gradient of the error committed in each training instance must be transferred from the last to the first layers, through a process known as "back-propagation". This is how neural networks learn from their mistakes [13]. However, when such networks are very deep, the proportion of the error that is transmitted from one layer to another is incrementally smaller, requiring an enormous amount of training time. In some instances, the gradient becomes 0, meaning that, from that point on, the network will learn nothing. Faced with this problem, Microsoft Research [14] presented as an alternative solution: the "Residual Network" (ResNet) and this study

employed the two most widely studied residual CNN models, ResNet50V2 and ResNet101V2 [15], and compare their results.

Based on an innovative architecture, ResNet utilize the idea of splitting a layer into two branches, where one branch does nothing to the signal and the other branch processes it as a typical layer would [16][17]. Figure 1 shows a fundamental block of the ResNet architecture where a hop connection is shown. In this block, there are two different paths. The first, is a vertical flow representing the conventional learning process of a neural network, where the input "x", the gradient, traverses the layers of the network during training. At each layer, "x" is multiplied by the layer weights and the activation function "ReLU" is applied to produce $F(x)$, and at the end of the layer we would have $H(x) = F(x)$ [10].

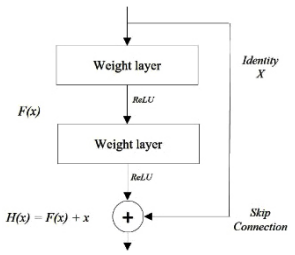


Fig. 1. Residual learning [10]

However, in ResNet there is a second alternate flow that occurs in parallel. In this second flow, on the right, a copy of the gradient input "x" is created and would be the identity matrix of "x", and this value is connected directly to the layer output, where it is added to $F(x)$, making $H(x)$ now $F(x) + x$ (Eq. 1). This prevents the gradient from disappearing after traversing layers, because the value of the input "x" is constantly being added. Furthermore, it ensures that a deep neural network will not have more error than a nondeep one [10].

$$H(x) = F(x) + x \tag{1}$$

Thus:

$$F(x) = H(x) - x \tag{2}$$

3 Related work

Different approaches have emerged in the literature that seek to detect cervical cancer by applying AI techniques, such as: CNNs [7], ML Algorithms [8], DL architectures [18], TL [6], among others. These techniques have been applied for four purposes: (i) to "classify" the level of cervical cancer [6], (ii) to "extract features" from an image representing the level of cancer [19], (iii) to "segment images", for example to separate those of the nucleus from those of the cytoplasm of a cancer cell [19], and

(iv) to "enhance" the images [20]. These studies have used images from different public datasets like SIPaKMeD [21], Herlev [22], Mendeley LBC [23] or from their own dataset, and a majority have analyzed images from conventional Pap smears. A minority have analyzed LBC Pap smears.

Regarding the research related to the "classification" of cervical cancer using conventional Pap smear images, [6] having classified three types of cervical cancer images (Type 1/Type 2/Type 3) by applying TL and 3 DL architectures (InceptionV3, ResNet50 and VGG19), using the Kaggle dataset [24] and considering 5,278 images, where their best result occured with the InceptionV3 model that achieved an accuracy = 96.1%. [7] proposed a CNN architecture with their own dataset consisting of 200 images, conducting the "classification" in two stages: (i) identification of two types of cancer cells, normal and abnormal and (ii) identification of seven sets of classes of cancer images, with three sets of normal cancer images and four sets of abnormal cancer images. They achieved an accuracy of 89%. [8] proposed a hybrid approach based on the construction of a CNN and the application of a support vector machine (SVM), utilizing three kernel functions, Linear, Polynomial and Gaussian Radial Basis Function (RBF), and found their best accuracy result of 93.67% with the SVM and linear functions.

Regarding the works related to "feature extractions" for classification using images of conventional Papanicolaou tests, [18][10] applied SVM to extract the features and ResNet101 to train the model in order to classify seven types of classes, producing results with 100.0% and 92.0% accuracy in the training and testing phases, respectively. Regarding the works that apply "image segmentation" to locate cervical cancer cells in liquid-based Papanicolaou test images, [9] proposed a network that applied a Series-parallel fusion network (SPFNet), and compared their results with other state-of-the-art object detection algorithms. [25] applied a Mask Regional CNN (Mask R-CNN) to the Thammasat University Hospital dataset in order to detect and analyze the cervical cell nucleus, looking for normal and abnormal nuclear features.

Alternatively, there are studies that applied "segmentation", "feature extraction" and "classification" of cervical cancer, using only images from conventional Papanicolaou tests and using the SIPaKMeD and Herlev datasets. Among them, studies performed "segmentation" of cancer cells to identify cell nuclei and cytoplasm using the iterative shape-based method [22], applying the K-Means algorithm [19] and the TE-DFO algorithm [26]. For feature "extraction" they applied the Random Forest (RF) [19][22][27] and the SqueezeNet model [26]. And for cancer cell "classification" they use two public datasets, SIPaKMeD and Herlev, and apply various classifiers: LD (linear discriminant) [15][12], SVM [19][22], k-nearest neighbor (KNN) [19][22][27], boosted trees and bagged three [19][22], Logistic regression and Decision tree [27] and Extreme learning machine [26].

Finally, [20] applied Dual Tree Complex Wavelet Transform to "enhance" the images of the Herlev dataset, and then classified them into four classes using the DL Resnet18 architecture.

4 Proposed approach

This study's automatic cervical cancer detection system is based on DL mechanisms and a TL methodology. The workflow (Figure 2) starts with the choice of the public dataset of LBC images of Pap smears from the Mendeley LBC [23]. An algorithm developed with Keras increased the number of images in this dataset to 2,376. Then, two pre-installed CNN models (ResNet50V2 and ResNet101V2) were selected. At this stage, hyperparameters were defined to adapt these models to the present study. The training and validation of these models was performed considering 80% of the dataset and the remaining 20% for testing, with metrics defined to evaluate the performance of each model and select one for the development of the expert system.

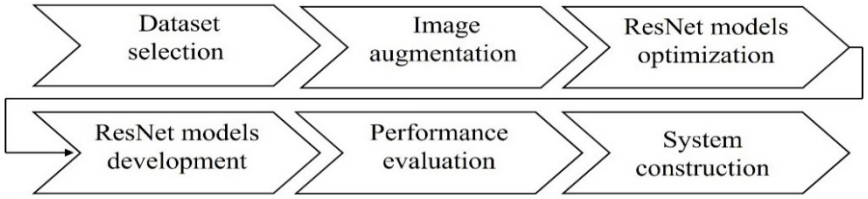


Fig. 2. Conceptualization of the proposed approach

4.1 Dataset

Public data from the Mendeley LBC dataset [23] containing 963 images of LBC specimens from Pap smears from a study of 460 patients were used (Table 1). These images represent the sub-categories of cervical lesions (malignant and pre-malignant) that follow the Bethesda System standard: Negative for Intraepithelial lesion or malignancy (NILM), Low-grade intraepithelial lesions (LSIL), High-grade intraepithelial lesions (HSIL) and Squamous Cell Carcinoma (SCC). Figure 3 shows the 4 types of images by type of carcinoma according to LBC Pap smears: NILM, LSIL, HSIL and SCC.

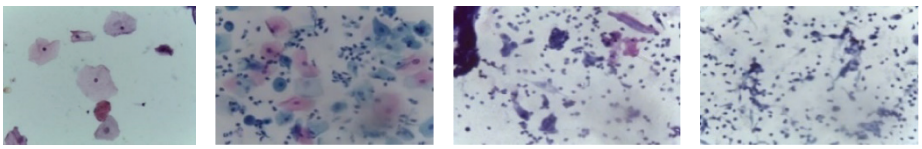


Fig. 3. Sample of the 4 types of images from the Mendeley LBC dataset [23]

4.2 Image augmentation

Since the Mendeley LBC dataset contains a relatively small quantity of images and the number of images per class is not balanced, data augmentation was performed with the *ImageDataGenerator* class in Keras [28]. Table 1 shows the number of images

per class, where it is observed that the "SCC" class of the Mendeley dataset has few images (74) compared to the rest.

Table 1. New dataset of LBC cervical cancer images

Category	Class	Dataset Mendeley LBC	Data Augmentation	New Dataset
NILM	1	613	35	648
LSIL	2	163	402	565
HSIL	3	113	500	613
SCC	4	74	476	550
		963	1,413	2,376

A python code was developed for the four image classes augmentation with *ImageDataGenerator*, the parameters used and their values are described in Table 2. After applying the data enhancement code to all the images in the four image categories, a new dataset of 2,376 images was generated.

Table 2. Parameters used with *ImageDataGenerator*

Parameters	Value	Description
rotation_range	45	Random image rotation between 0 and 45 degrees
width_shift_range	0.2	Image displacement along *X*-axis by 20%
height_shift_range	0.2	Image displacement along *Y*-axis by 20%
shear_range	0.2	Shear the image by 20%
zoom_range	0.2	Zoom-in and zoom-out by 20%.
horizontal_flip	true	For mirror reflection
fill_mode	reflect	When the image is shifted by 20%, there is some space left over. The "reflect" method will fill the area with the image reflection.

4.3 ResNet model optimization

As this study used the TL [6] methodology, previously trained weights from *ImageNet* were taken for the training of both models [11]. For this, the layers corresponding to the classification were disabled and the ImageNet weights were loaded to the rest of the layers. This intends to only train those layers used in the classification, so that they learn to detect the four classes of lesions and reduce the time spent on training. The training was carried out using the 2,376 images of the dataset, using different hyperparameter adjustment techniques, varying the learning rate, the number of epochs, and the activation functions, while, as will be explained below, adding new layers to the models.

The ResNet50V2 model was modified by disabling the *Average Pooling* operation and all the layers corresponding to the classification section (Figure 4). This modification looks to add more layers that fit the classification of the type of image considered for this study. As shown, the first layer added to the network is the "Flatten" layer that resizes of the layers by flattening them. Subsequently, 3 more layers were added:

"Dense" with *ReLU* activation function, in order to avoid the problem of overfitting, "Dropout" and "Gaussian Noise" which have a regularizing function that occurs during training. Additionally, it was observed that the dimensionality of the outputs decreases with the use of the "Dense" layer. Finally, a "Dense" layer with the *softmax* activation function was added, used for the output of the four neurons that correspond to the four classes of lesion images.

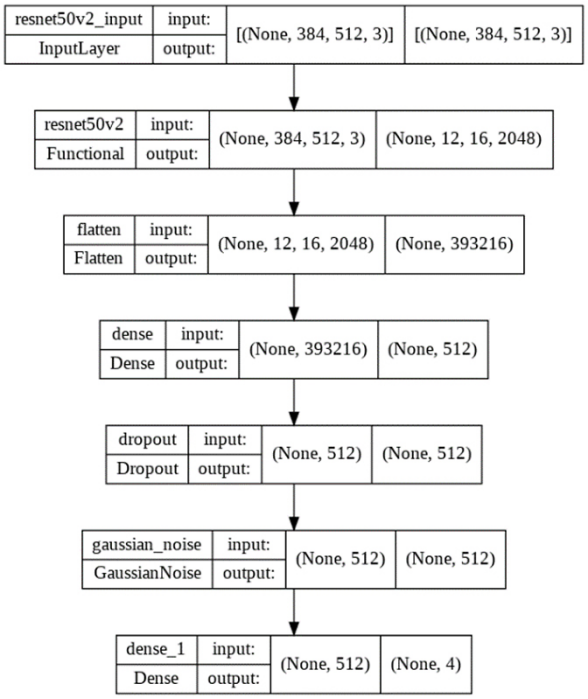


Fig. 4. Matching the ResNet50V2 architecture with tensorflow

For compilation of the ResNet50V2 and ResNet101V2 models the following parameters were used. The "optimizer" parameter employed is of the *Adam* type, designed to adjust the learning rate to the network requirements. The "learning rate" of the model is used to control the learning steps (1e-4). For "loss function" a *sparse_categorical_crossentropy* was determined best to classify the four classes of images. "Performance metric" was set to accuracy. A *batch_size* of 256 samples was used, that had a total of 50 epochs as a maximum for the whole training, and *Early Stopping* with *Patience* equal to 12 so that the training stops at the most optimal number of epochs, and where the highest accuracy is found, in order to conserve resources and avoid overfitting of the model.

Figure 5 shows the Python code developed to train the ResNet50V2 model, where the first step looked to initialize the variables dimensions, number of images and learning rate.

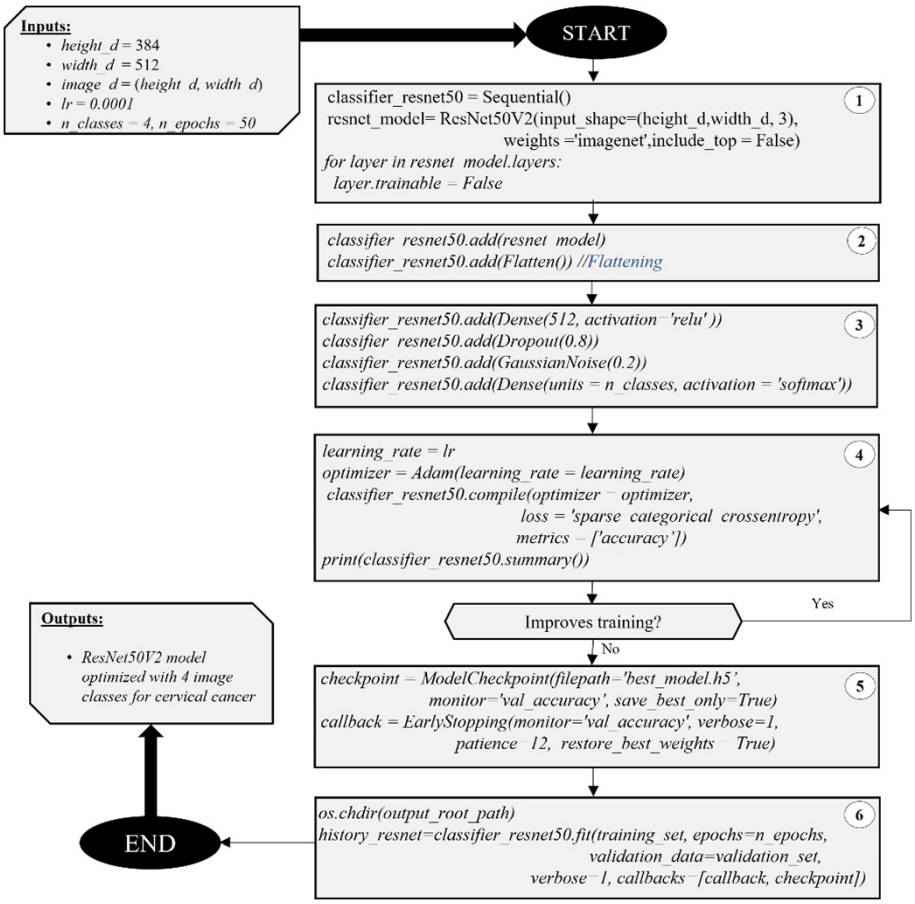


Fig. 5. ResNet50V2 training steps

The initialization of the CNN was performed, where the variable *resnet_model* was instantiated with the pre-trained ResNet50V2 model, the *Imagenet* weights were applied, and the last fully connected layers were deactivated. In steps 2 and 3, the model was built by adding additional layers and activation functions, in order to classify the four types of images. In step 4, the model was compiled with the defined parameters and included a printed summary of the results. At this point a determination is made as to whether the training is improving. If there is no improvement, *Early Stopping* was applied as step 5, so that the training stops at the most optimal number of epochs. Finally, the model was trained with step 6. The ResNet101V2 model training process was executed the same way. The names of the models were changed to ResNet50V2* and ResNet101V2*, with the serving differentiate this study's optimized models from the original CNN models. The experiment was conducted in a *Keras* 2.2.4 environment with *Tensorflow* 2.0 and Theano 1.0 backend. Due to the extensive CPU capacities for the training of both models, the services of *Google Colab* were used.

5 Results and discussion

5.1 Evaluation metrics

Different classification metrics evaluate the performance of the pre-trained models such as: *Precision* (Eq. 3), *Recall* (Eq. 4), *F1 Score* (Eq. 5) and *Accuracy* (Eq. 6).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN(Number\ of\ Correct\ Classified\ Image)}{TP + TN + FP + FN\ (Number\ of\ All\ Images)} \tag{6}$$

Where,

- *TP* (True positives): Images of cervical dysplasia correctly classified.
- *TN* (True negatives): Images of cervical non-dysplasia correctly classified.
- *FP* (False positives): Images of cervical dysplasia incorrectly classified.
- *FN* (False negatives): Images of cervical non-dysplasia incorrectly classified.

The "precision" or true negative rate (*TN* rate) is the proportion of true negatives that are expected to be negative. The recall or true positive rate (*TP* rate) is the proportion of true positives predicted to be positive. The "F1score" helps compare the performance of the two classifiers, representing a widely used measure of balance calculation for unbalanced data sets. The "accuracy" expresses the overall efficiency in image classification accuracy [29][30].

5.2 Training results

As the Pareto principle was used to separate the training and validation images, 80% of the images were used for training and validation and the remaining 20% were used for testing. The results for "accuracy" in the training and validation for ResNet50V2* (Figure 6a) presented an average of 100% and 95.78% in the training and validation phases, respectively. Furthermore, it is observed that from epoch 8 onwards, there is uniformity in the validation stage. The "loss" in training and validation decreased slightly as the epochs progressed, reaching the end of training with a loss equal to zero (Figure 6b). Similarly, in the ResNet101V2* model, from epoch 15, there is a uniformity in the validation (Figure 6c). As well, the "loss" in the training and validation in the ResNet101V2* model decreased slightly as the epochs progressed, reaching the end of training with a "loss" equal to zero (Figure 6d).

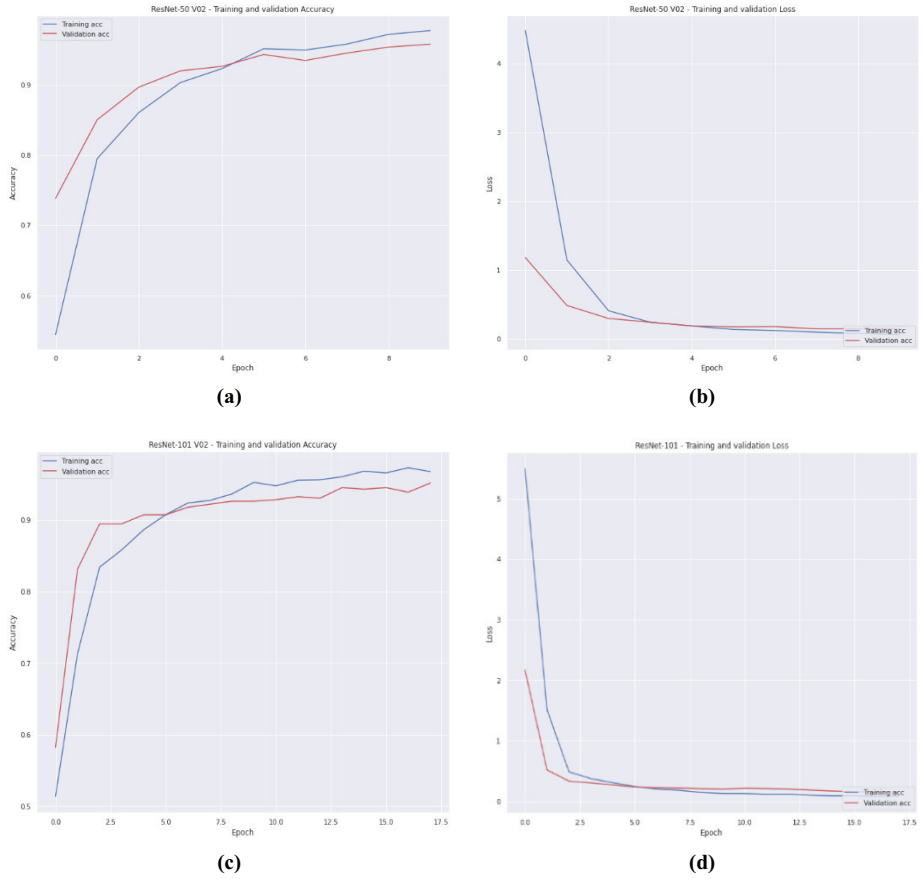
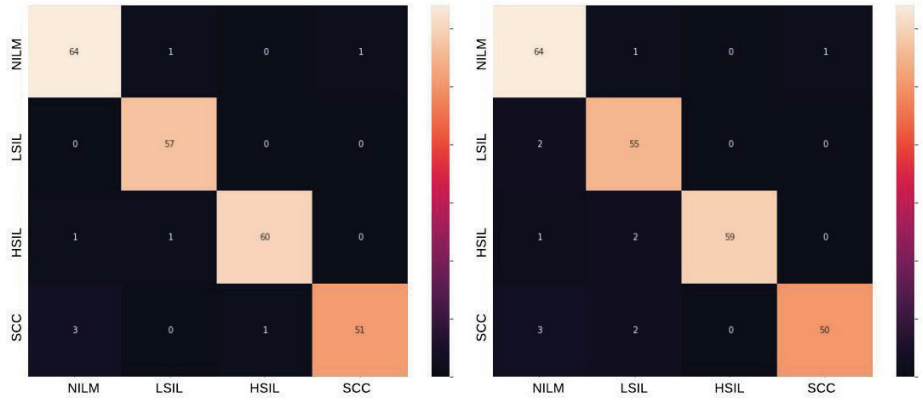


Fig. 6. Accuracy and loss in training and validation with RestNet50V2* and ResNet101V2*

Testing was performed with the four image classes and the confusion matrices of the ResNet50V2* and ResNet101V2* (Figure 7) model were obtained. Table 3 shows a summary of the number of correct and incorrect predictions of the four types of images made by each model.



ResNet50V2* **ResNet101V2***

Fig. 7. Confusion matrix

Table 3. Number of correct and incorrect predictions with ResNet50V2* and ResNet101V2*

Network	Image Type	Correct Predictions	Incorrect Predictions	Total
ResNet50V2*	NILM	64	2	66
	LSIL	57	0	57
	HSIL	60	2	62
	SCC	51	4	55
ResNet101V2*	NILM	64	2	66
	LSIL	55	2	57
	HSIL	59	3	62
	SCC	50	5	55

Based on the data obtained from the confusion matrices, Eq. (3), Eq. (4), Eq. (5) and Eq. (6) were applied to generate the respective metrics of Resnet50V2* and ResNet101V2* (see Figure 8).

	precision	recall	f1-score	support
0	0.94	0.97	0.96	66
1	0.97	1.00	0.98	57
2	0.98	0.97	0.98	62
3	0.98	0.93	0.95	55
accuracy			0.97	240
macro avg	0.97	0.97	0.97	240
weighted avg	0.97	0.97	0.97	240

ResNet50V2*

	precision	recall	f1-score	support
0	0.91	0.97	0.94	66
1	0.92	0.96	0.94	57
2	1.00	0.95	0.98	62
3	0.98	0.91	0.94	55
accuracy			0.95	240
macro avg	0.95	0.95	0.95	240
weighted avg	0.95	0.95	0.95	240

ResNet101V2*

Fig. 8. Parameters of evaluation with Python

Table 4 shows the summary of the metrics obtained by each model. The optimized ResNet50V2* model showed the highest accuracy of 0.98 for the HSIL (High-grade intraepithelial lesions) and SCC (Squamous cell carcinoma) images. Its best true

positive rate of 1.0 was found with the LSIL (Low-grade intraepithelial lesions) images. A better balance between accuracy and Recall, 0.98, was found with the and HSIL images. The optimized ResNet101V2* model achieved the best accuracy, 1.00, for the HSIL images. The best true positive rate, 0.97, was found for the NILM (Negative for Intraepithelial lesion or malignancy) images. And the best balance between accuracy and Recall, 0.98, was found for the HSIL images. The two models showed good accuracy in classifying the four types of cervical cancer images, with better accuracy, 0.97 found with the optimized ResNet50V2* model.

Table 4. Summary of metrics obtained with the optimized models by image type

Image Type	ResNet50V2*			ResNet101V2*		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
NILM	0.94	0.97	0.96	0.91	0.97	0.94
LSIL	0.97	1.00	0.98	0.92	0.96	0.94
HSIL	0.98	0.97	0.98	1.00	0.95	0.98
SCC	0.98	0.93	0.95	0.98	0.91	0.94
Accuracy	0.97			0.95		

5.3 Comparison with other studies

Table 5 compares this study's two optimized models with recent literature approaches for artificial intelligence image classification for different classes of cervical cancer. Almost all the studies focused on cervical cancer image classification based on conventional smears, with one study found [25] that focused on LBC, but had the least amount of image samples, only detected two classes, and achieved a sub-.90 accuracy. In terms of sample size, the study with the most image samples [6], was based on conventional smears, and designed to detect three classes, ultimately achieving an accuracy similar to this study.

Table 5. Existing studies ordered by test and then accuracy

AI Methods	Image Type	Accuracy	Classes	Samples
ResNet50V2*	LBC	0.97	4	2,376
ResNet101V2*	LBC	0.95	4	2,376
Mask R-CNN [25]	LBC	0.89	2	178
InceptionV3 [6]	Conventional smears	0.96	3	5,278
RF [27]	Conventional smears	0.95	7	917
CNN and SVM [8]	Conventional smears	0.94	2	652
ML algorithms [22]	Conventional smears	0.94	5	917
CNN [7]	Conventional smears	0.89	3	200
ML algorithms [19]	Conventional smears	0.82	7	917

5.4 Expert system development

For the development of the expert system, the optimized ResNet model with the best performance was employed: ResNet50V2*. Figure 9 shows the integrated architecture of the proposed web system, with the technologies used, user types (administrator, doctor and director) as well as the functionality of the system consisting of management, diagnostic and report modules.

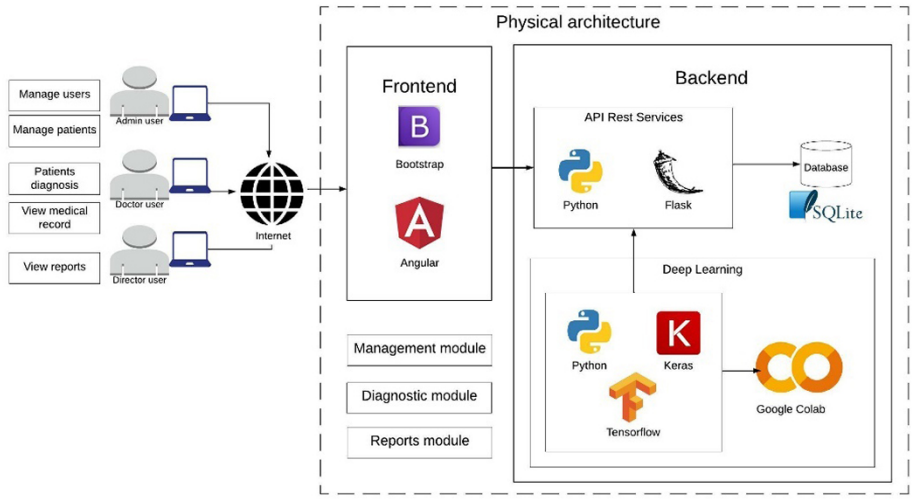


Fig. 9. Approach integration architecture

For the validation of the expert system, images from six patients from a specialized cancer institution located in Peru were analyzed. The data sample consisted of six LBC images of Papanicolaou tests (Figure 10), with diagnoses performed by a cytologist-pathologist expert. The expert diagnosis of the images was HSIL, and the system was validated as it also generated the same HSIL diagnosis for the same six images.

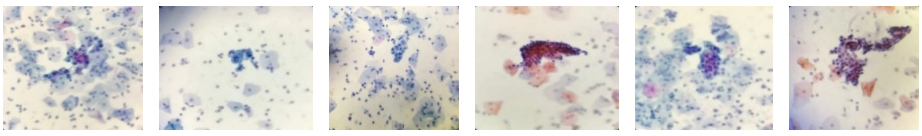


Image 1 **Image 2** **Image 3** **Image 4** **Image 5** **Image 6**

Fig. 10. Sample of Papanicolaou LBC images of 6 patients

6 Conclusions

An approach to build an automatic cervical cancer detection system based on image processing of LBC Pap smear tests was proposed. The proposal was developed based on DL mechanisms and TL methodology. A dataset was constructed based on a public Mendeley LBC dataset, divided into 4 classes of precancerous and cancerous cervical

cancer lesions: NILM, LSIL, HSIL and SCC. Based on this dataset, data augmentation algorithms were developed to obtain a total of 2,376 images. The TL methodology was applied by building 2 pre-trained DL models (ResNet50V2* and ResNet101V2*) using Tensorflow and Keras tools, and optimizing each model by defining hyperparameters. The application of the *Early Stopping* hyperparameter helped reduce execution time and resource consumption, while preventing overfitting of the models. For the evaluation of each optimized model, 4 metrics were adopted: *Precision, Recall, F1-Score* and *Accuracy*. Regarding the performance of the ResNet50V2* model, it achieved better accuracy for the HSIL and SCC images and a better true positive rate for the LSIL images. The ResNet101V2* model obtained better precision for the HSIL images and a better balance between *presision* and *Recall* for the HSIL images. ResNet50V2* obtained best performance overall in the study and when compared with previous research. A system based on the ResNet50V2* model was built and validated, demonstrating a helpful automated system to aide doctors in diagnosing cervical cancer.

7 Acknowledgment

We would like to thank the Vice-Rectorate of Research of the National University of San Marcos for funding the project.

8 References

[1] MINSA, "Guía de práctica clínica para la prevención y manejo del cáncerde cuello uterino," *Ministerio de Salud*, 2017, http://bvs.minsa.gob.pe/local/MINSA/4146.pdf (accessed Oct. 05, 2021).

[2] H. Zur Hausen, "Papillomaviruses and cancer: from basic studies to clinical application," *Nat. Rev. Cancer*, vol. 2, no. 5, pp. 342–350, 2002, https://doi.org/10.1038/nrc798

[3] G. González-Martínez, "George N. Papanicolaou (1883–1962)," *Rev. Obstet. Ginecol. Venez.*, vol. 65, no. 1, 2005, [Online]. Available: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0048-77322005000100008

[4] M. Puerto de Amaya, P. Moreno-Acosta, and M. Mora, "Conventional and liquid-based cytology in a shared sample of cervical shots," *Repert. Med. y Cirugía*, vol. 24, no. 1, pp. 41–46, 2015, https://doi.org/10.31260/RepertMedCir.v24.n1.2015.652

[5] GLOBOCAN, "Estimated age-standardized incidence and mortality rates (World) in 2020, worldwide, females, all ages," 2020, https://gco.iarc.fr/today

[6] S. Dhawan, K. Singh, and M. Arora, "Cervix image classification for prognosis of cervical cancer using deep neural network with transfer learning," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 7, no. 27, pp. 1–9, 2021, https://doi.org/10.4108/eai.12-4-2021.169183

[7] N. S. Manasa Ungrapalli and A. N. Myna, "Classification of pap smear images for cervical cancer using convolutional neural network," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 2801–2807, 2019, https://doi.org/10.35940/ijitee.J1226.119119

[8] J. E. Aurelia, Z. Rustam, and I. Wirasati, "Cervical cancer classification using convolutional neural network-support vector machine," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 19, no. 5, pp. 1605–1611, 2021, https://doi.org/10.12928/telkomnika.v19i5.20406

[9] M. Xia, G. Zhang, C. Mu, B. Guan, and M. Wang, "Cervical cancer cell detection based on deep convolutional neural network," *Chinese Control Conf. CCC*, vol. 2020-July, pp. 6527–6532, 2020, https://doi.org/10.23919/CCC50068.2020.9188454

[10] M. Nash, *Deep Learning, Computer-Aided Radiography Reading for Tuberculosis: A Diagnostic Accuracy Study from a Tertiary Hospital in India. Montreal, Canada:* McGill University.

[11] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *IJCV*, vol. 115, pp. 211–252, 2015, [Online]. Available: https://link.springer.com/article/10.1007/s11263-015-0816-y?sa_campaign=email/event/articleAuthor/onlineFirst; https://doi.org/10.1007/s11263-015-0816-y

[12] A. Errabih et al., "Identifying retinal diseases on OCT image based on deep learning," *Int. J. Online Biomed. Eng*, vol. 18, no. 15, pp. 141–159, 2022, https://doi.org/10.3991/ijoe.v18i15.33639

[13] S. Showkat and S. Qureshi, "Efficacy of transfer learning-based ResNet models in chest X-ray image classification for detecting COVID-19 pneumonia," *Chemom. Intell. Lab. Syst.*, vol. 224, no. February, p. 104534, 2022, https://doi.org/10.1016/j.chemolab.2022.104534

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, https://ieeexplore.ieee.org/document/7780459; https://doi.org/10.1109/CVPR.2016.90

[15] Keras, "ResNet and ResNetV2." https://keras.io/api/applications/resnet/ (accessed Dec. 07, 2022).

[16] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing*. Boca Raton: CRC Press, 2017, https://doi.org/10.4324/9781315154282

[17] S. N. M. Safuan, M. R. M. Tomari, and W. N. W. Zakaria, "Cross validation analysis of convolutional neural network variants with various white blood cells datasets for the classification task," *Int. J. Online Biomed. Eng.*, vol. 18, no. 2, pp. 123–140, 2022, https://doi.org/10.3991/ijoe.v18i02.27321

[18] H. Alquran et al., "Cervical cancer classification using combined machine learning and deep learning approach," *Comput. Mater. Contin.*, vol. 72, no. 3, pp. 5117–5134, 2022, https://doi.org/10.32604/cmc.2022.025692

[19] K. P. Win, Y. Kitjaidure, M. P. Paing, and K. Hamamoto, "Cervical cancer detection and classification from pap smear images," *ACM Int. Conf. Proceeding Ser.*, pp. 47–54, 2019, https://doi.org/10.1145/3366174.3366178

[20] T. N. Vijayanand Sellamuthu Palanisamy, Rajiv Kannan Athiappan, "Pap smear based cervical cancer detection using residual neural networks deep learning architecture," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 4, 2022, https://doi.org/10.1002/cpe.6608

[21] Q. W. M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, "Deepcervix: a the classification of cervical cells using hybrid deep feature fusion techniques," *Comput. Biol. Med.*, vol. 136, 2021, https://doi.org/10.1016/j.compbiomed.2021.104649

[22] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. M. Aung, "Computer-assisted screening for cervical cancer using digital image processing of pap smear images," *Appl. Sci.*, vol. 10, no. 5, 2020, https://doi.org/10.3390/app10051800

[23] E. Hussain, "Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer." Mendeley, 2019, https://doi.org/10.17632/zddtpgzv63.4

[24] Kaggle, "Kaggle Code." https://www.kaggle.com/code

[25] N. Sompawong et al., "Automated pap smear cervical cancer screening using deep learning," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 7044–7048, 2019, https://doi.org/10.1109/EMBC.2019.8856369

[26] M. I. Waly, M. Y. Sikkandar, M. A. Aboamer, S. Kadry, and O. Thinnukool, "Optimal deep convolution neural network for cervical cancer diagnosis model," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 3297–3309, 2022, https://doi.org/10.32604/cmc.2022.020713

[27] H. Bandyopadhyay and M. Nasipuri, "Segmentation of pap smear images for cervical cancer detection," *2020 IEEE Calcutta Conf. CALCON 2020—Proc.*, pp. 30–33, 2020, https://doi.org/10.1109/CALCON49167.2020.9106484

[28] TensorFlow V2.11.0, "keras.preprocessing.image.ImageDataGenerator." https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (accessed Oct. 10, 2022).

[29] T. R. Hagedoorn and G. Spanakis, "Massive open online courses temporal profiling for dropout prediction," *Proc. 2017 IEEE 29th Int. Conf. Tools with Artif. Intell.*, pp. 231–238, 2017, [Online]. Available: https://ieeexplore.ieee.org/document/8371948/metrics#metrics

[30] S. M. Naser, Y. H. Ali, and D. A. J. Obe, "Hybrid cyber-security model for attacks detection based on deep and machine learning," *Int. J. Online Biomed. Eng.*, vol. 18, no. 11, pp. 17–30, 2022, https://doi.org/10.3991/ijoe.v18i11.33563

9 Authors

Lenis Wong. Is a professor and researcher of Software Engineering and Information Systems Engineering at the National University of San Marcos and Peruvian University of Applied Sciences, Peru. She holds a PhD in Systems Engineering and Computer Science. MSc. in Systems and Computer Engineering with mention in Software Engineering. She is a member of the AI research group and has carried out different multidisciplinary projects where Artificial Intelligence, Software Engineering and Information Systems Engineering have been applied in education, health, medicine and healthcare. She has published several international peer-reviewed scientific articles in different multidisciplinary areas such as: ML, DL, IoT, e-Health, Software Engineering, Requirements Engineering, Cloud Computing, E-Learning, Gamification, Cyberattacks, Natural Language Processing, Networks and Blockchain Technologies (email: lwongp@unmsm.edu.pe).

Andrés Ccopa. Is a software engineer and former student of Dr. Lenis Wong's thesis course at the Faculty of Systems and Computer Engineering of the National University of San Marcos, Peru. His area of development includes Backend development specialized in Java language, mobile development with Android and artificial intelligence, especially in neural networks (email: andres.ccopa@unmsm.edu.pe).

Elmer Diaz. Is a software engineer and former student of Dr. Lenis Wong's thesis course at the Faculty of Systems and Computer Engineering of the National University of San Marcos, Peru. His area of development includes backend development specialized in Java language, mobile development with Android and artificial intelligence, especially in neural networks (email: elmer.diaz2@unmsm.edu.pe).

Sergio Valcarcel. Is currently a PhD candidate in Systems Engineering at the National University of San Marcos. He holds a Master's degree in Systems Engineering and Computer Science. Assistant professor at the Universidad Nacional Mayor de San Marcos, in undergraduate and graduate courses. He is a member of the AI research group, has published 2 books in this line of research and is a reviewer of scientific articles in AI at the Instituto Tecnológico de la Producción. He is experienced in the

implementation of information technology projects applying AI concepts at the Servicio Nacional de Sanidad Agraria (email: svalcarcela@unmsm.edu.pe).

David Mauricio. Is a professor and researcher of Information Systems Engineering at the National University of San Marcos, Peru. Received the M.S. degree in applied mathematics and the Ph.D. degree in systems and computing engineering from the Federal University of Rio de Janeiro, Brazil. He was a Professor with the North State University Fluminense, Brazil, from 1994 to 1998. Since 1998, he has been a Professor with the National University of San Marcos. His research interests include mathematical programming, AI, software engineering, and entrepreneurship (email: dmauricios@unmsm.edu.pe).

Vladimir Villoslada. Is currently a Gynecologic Oncologist of the Department of Gynecologic Surgery of the National Institute of Neoplastic Diseases. Medical surgeon graduated from the National University of Cajamarca, specialist in Gynecology and Obstetrics from the National University of Trujillo and sub-specialty in Oncological Gynecology from the Peruvian University Cayetano Heredia, candidate for a master's degree in Epidemiological research sciences. His areas of interest are epidemiology, biostatistics, data science, machine learning applied to cancer. Postgraduate professor at the University Cayetano Heredia and associate researcher at the Faculty of Medicine of the National University of Cajamarca (email: vladimirvilloslada@gmail.com).

Article submitted 2022-12-16. Resubmitted 2023-01-25. Final acceptance 2023-01-27. Final version published as submitted by the authors.