Image Quality Assessment of Cytology Images using Deep Learning

Tomé Albuquerque[1,2], Maria João M. Vasconcelos[3], Jaime S. Cardoso[1,2]

[1] Faculty of Engineering of the University of Porto, Porto, Portugal; [2] INESC TEC, Porto, Portugal ; [3] Fraunhofer Portugal AICOS, Porto, Portugal

1 Introduction

Medical **image quality assessment** plays an important role not only in the design and manufacturing processes of image acquisition but also in the **optimization of decision systems**. **Cervical cancer** remains the fourth most common cause of cancer death in women worldwide (Figure 1). Despite the outburst of recent scientific advances to find an effective treatment, there is no effective method, especially when diagnosed in an advanced stage. However, **screening tests** such as **cytology** or colposcopy, have been responsible for a **strong decrease** in cervical cancer deaths. Cytology microscope images need high-level microscopic magnification for a consistent characterization, but it is necessary to **preserve** an appropriate **image quality** [1-3]. In this work, some approaches for non-reference image quality assessment (NR-IQA) are presented using different convolutional neural network (CNN) based models fine-tuned to predict the quality score value of several images. The first models use three different architectures (**VGG16, MobileNet, and ResNet50**) to predict the image quality of a **blood cells microscopy dataset**, labelled in a **multiclass problem** (4 classes). Due to the lack of annotated cytology dataset regarding image quality, it was used as reference a microscopic blood cell sample dataset. After selecting the best model, the weights of the best model were used to initiate the train of a new model to **classify IQA** on a **new dataset** created in this work with **cytology microscope images**. The classification of IQA on the cytology dataset is done on a binary problem (bad quality vs good quality).

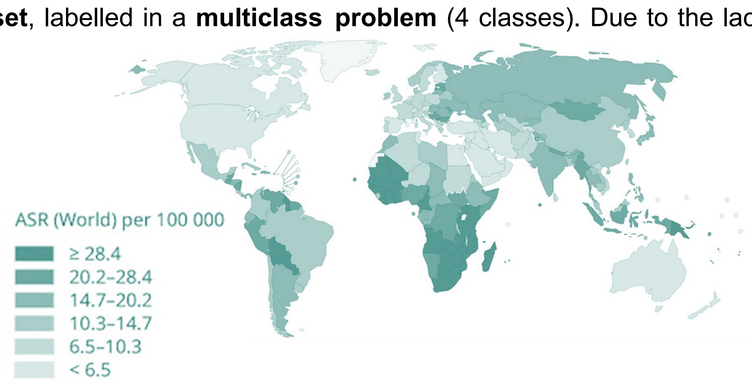

Figure 1 — Age standardized world incidence rates of cervical cancer (WHO, 2018).

2 Methods

Dataset

Two different datasets were used in the train of IQA models. The microscope slide preparations of the **first dataset**, from **blood samples**, were obtained in Centro Hospitalar de São João (Porto), and digitalized by Fraunhofer AICOS (FhP-AICOS) within the scope of MpDS project using µsmartscope (Figure 2). The blood cells dataset contain a total of 1854 images divided into **4** different **classes** according to quality score of the image, which can be **bad, fair, good, or excellent quality**. The **second dataset** was collected in Hospital Professor Doutor Fernando Fonseca (Lisbon) within the scope of CLARE project using an updated version of µsmartscope, adapted to acquire samples with 400 times magnification the images were also acquired as the first dataset by FhP-AICOS using µsmartscope. This second dataset consists of 4088 images of **microscope pap smear slide preparations (liquid-based citology samples)**, in which 817 are reference images and 3271 are images generated from the reference images with four different types of distortions. Thus, for every image of reference four new images were created with the distortions mentioned before as showed in Figure 3. This new cytology dataset is divided in **2** different **classes**, **distorted images** (bad quality images), and **reference images** (good quality images).

Figure 2 — µsmartscope by Fraunhofer AICOS (FhP-AICOS).

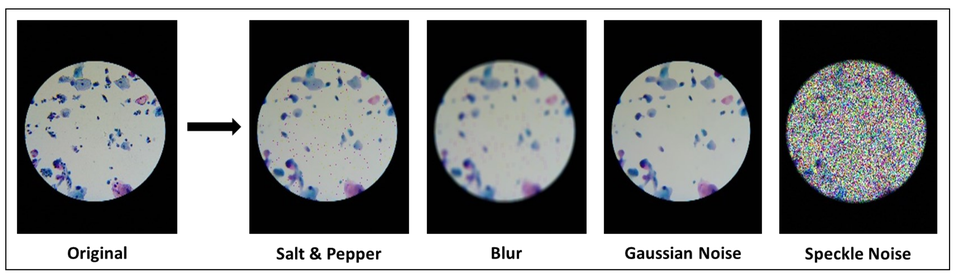

Figure 3 — Example of the four different types distortions applied on the original images of cytology dataset.

Data Pre-Processing

1. Split datasets in train, validation, and test subsets (60/20/20%) for 5 different k-folds;
2. Resize images to 224 x 224 pixels;
3. Normalize from the pixel values range of 0—255 to 0—1;
4. Data augmentation during the training including image rotation with a range of 90º, width and height shift (10%), horizontal flip, zoom (in or out) (10%).

Experimental Details

Two different models were trained and tested in this work (Figure 4). The first model 1 was tested with three different CNN architectures. All the architectures were directly fed with the blood cells and cytology images. The CNN architecture that achieves the best results in the IQA of blood cells dataset was used on model 2 for cytology IQA. The tested CNN architectures are the following: VGG16, MobileNet and ResNet50. For multiclass IQA of the blood cells, it was implemented the model 1 with the 3 different convolutional architectures mentioned above. Model 2 follows the same pattern of model 1, however, it is a CNN model for binary classification. This model classifies the image quality of cytology cells in two different classes. The pre-trained weights of model 1 are used to initiate the train of model 2 through transfer learning to increase the performance and diminish training time.

The hyperparameters were adjusted just for model 1 with the grid search where the hyperparameters tuned were the learning rate (LR ∈ [0.001; 0.0001]), the batch size (BS ∈ [16; 32]), and the dropout rate (Dt ∈[0.2; 0.5]).

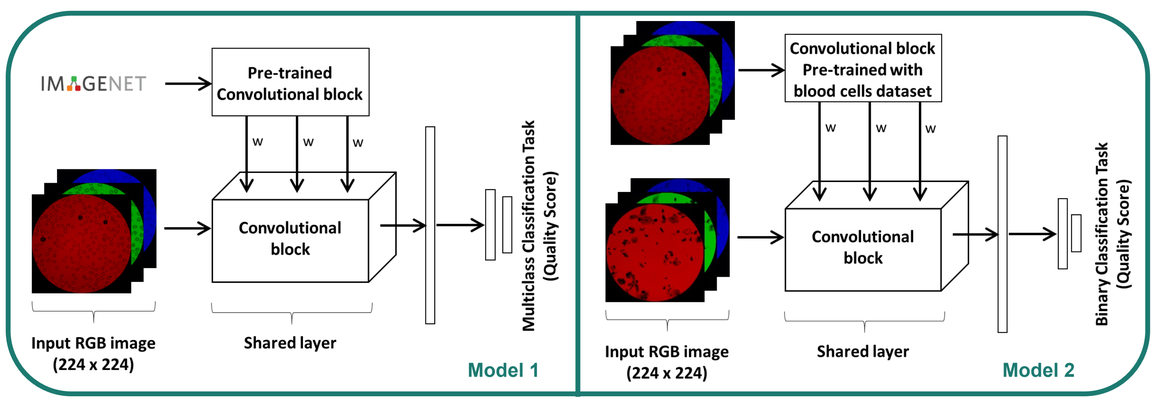

Figure 4 — Schematic diagram of model 1 and 2 for multiclass and binary classification task, respectively.

3 Results

The best combination of the hyperparameters found for model 1 after a fine-tuning was LR of 0.0001, BS of 32, and Dt of 0.2.The best results after the grid search for the first model (model 1 - multiclass) are represented in Table 1. For the multiclass problem, VGG16 achieved the best results with higher values in all the presented metrics when compared with MobileNet and ResNet50.

Table 1 — Results of model 1 for multiclass classification task of IQA in blood cells test subset.

	Accuracy (%)	Precision (%)	recall (%)	AUC (%)
MobileNet	76.05± 1.74	76.34 ± 2.75	76.05 ± 1.74	86.85 ± 1.39
VGG16	**78.91 ± 1.97**	**79.16 ± 2.17**	**78.91 ± 1.97**	**87.64 ± 1.82**
ResNet50	76.97 ± 2.39	77.29 ± 2.67	76.97 ± 2.39	83.06 ± 3.38

To confirm if the resize of the images to 224 x 224 did not contribute to loss of information an additional model trained with resized images but with bigger dimensions (512 x 512) was created and tested. Due to the imbalance in the number of images per class in blood cells dataset classes a new model was done. Thus, to guarantee that the imbalance does not interfere with the performance of the models a different model was tested were the minority classes were oversampled to get the same number of images of the class with a high number of images. The results for these different approaches using different images sizes and with or without oversampling shows no upgrades in the metrics. The model 1 trained with VGG16 as the convolutional block and using as input 224 x 244 resized images achieved the best performance. Thus, the pre-trained weights of this model 1 will be used by model 2 which will be only trained with a VGG16 architecture for binary IQA of cytology images. The last layer was discarded. The results of model 2 are presented in the following table 2.

Table 2 — Results of model 2 using VGG16 architecture to assess image quality of pap smear cells (cytology dataset).

	Accuracy (%)	Precision (%)	recall (%)	AUC (%)
VGG16	99.49 ± 0.72	99.50 ± 0.70	99.49 ± 0.72	99.96 ± 0.07

VGG16 in the multiclass quality assessment of blood cells images achieved the best performance in all metrics. The model proposed for classification of the pap smear images quality achieved a good performance for the binary image quality classification task with 99.49% of accuracy. This algorithm classifies the images according to the presence or absence of distortions, this way the classifier is focused on low-level notions of quality. To classify the images taking in account semantic complex concepts it is necessary to provide more information to the model.

4 Conclusions and Future Work

→The use of IQA in biomedical applications is essential to help in optimization and improvement not only in image processing techniques but also on diagnostic algorithms. Nevertheless, the use of IQA methods in medical applications is very limited to low notions of quality, such as distortions or noise on the images. It is essential to collect more data and semantic information about image quality to build more robust and accurate algorithms to assess quality.

→In this work it was demonstrated the good performance of deep learning algorithms using CNN for NR-IQA in biomedical databases.

→For future works, since a screening system is expected to be able to avoid misclassifying, artifacts must be added to the synthetic dataset of cytology to test the capacity of the CNN classifiers to detect that. In the future, it may be also interesting, add noise only in part of the image, and train with these examples. After that, analyze the activation maps and see if the explanation is consistent with the spatial placement of the noise.

Acknowledgements:

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project ``POCI-01-0145-FEDER-028857".

References:

[1] Ewert Bengtsson and Patrik Malm. Screening for cervical cancer using automated analysis of pap-smears.Computational and mathematical methods in medicine, 2014:842037, 03 2014.

[2] Zhi Lu, Gustavo Carneiro, Andrew Bradley, Daniela Ushizima, Ma-soud S. Nosrati, Andrea Bianchi, Cláudia Carneiro, and GhassanHamarneh. Evaluation of three algorithms for the segmentation of overlapping cervical cells. IEEE Journal of Biomedical and HealthInformatics, 21:1–1, 01 2016.

[3] Teresa Conceição, Cristiana Braga, Luís Rosado, and Maria Vasconcelos. A review of computational methods for cervical cells segmentation and abnormality classification.International Journal of Molecular Sciences, 20:5114, 10 2019.