Cervical Cancer Detection with the Papanicolaou Smear with Shallow and Deep Classifiers

Arthur S. Quadros[1], **Quézia P. Silva**[1], **Sarah S. Magalhães**[1]

[1]Institute of Exacts Sciences and Informatics
Pontifical Catholic University of Minas Gerais (PUC Minas)
30.535-901 – Belo Horizonte – MG – Brazil

{aquadros,qsilva,sarah.magalhaes.1280966}@sga.pucminas.br

Abstract. *This paper investigates the efficacy of shallow and deep learning techniques for cervical cancer detection using Pap smear images. We employ Haralick descriptors of six co-occurrence matrices as features for a Support Vector Machine (SVM) classifier and utilize the entire image for the deep learning model EfficientNetB1. The results demonstrate the superiority of EfficientNetB1, achieving an accuracy of 85% and 65% in the binary and multi-class (6 classes) classification tasks, respectively. Conversely, the SVM classifiers achieved lower accuracies of 64% and 47% for the binary and multi-class problems. These findings suggest that deep learning approaches hold promise for improved cervical cancer detection.*

1. Introduction

The Papanicolaou test, commonly known as the Pap smear, is a critical procedure in the early detection of cervical cancer. By collecting cells from the cervix and examining them under a microscope, this test can identify potentially precancerous and cancerous processes [Koss 1989, Naucler et al. 2007, Clarke and Anderson 1979]. The integration of AI for this purpose leverages the analysis of medical images, enhancing accuracy and speed in cancer detection. AI algorithms can learn from vast datasets to recognize patterns indicative of cancer, making the Pap smear an even more potent tool in combating this disease.

In this study, we employed a dual approach to classify Pap smear images from the dataset of [Rezende et al. 2021]: a shallow classifier using Haralick Descriptors – entropy, homogeneity, contrast – derived from multiple cooccurrence matrices, and a deep classifier analyzing pixel matrices from colored images via EfficientNetB1, specified in section 3. These methods were applied to both binary and multi-class contexts. Data augmentation increased our dataset fourfold with rotated images at 0, 90, 180, and 270 degree angles. We assessed our models' performance using accuracy and confusion metrics, as well as loss specifically for EfficientNetB1.

The implementations itself were made in Python with the libraries Scikit-Learn for the use of SVM, Tensorflow for the use of EfficientNetB1, OpenCV and Pillow for preprocessing and manipulating images, numpy for calculating the Haralick Descriptors, and Tkinter for creating a simple graphical interface.

2. Related Works

Similar studies use shallow and deep learning techniques in order to recognize patterns that indicate cancer. [Chen et al. 2023] achieved over 90% accuracy in 3-class classifers

using EfficientNetB0 deep classifiers, [Rastogi et al. 2023] achieved 94% accuracy in benchmarking datasets for binary classification of "normal" and "cancerigenous" cells using EfficientNetB7 deep classifiers. [Zhang and Liu 2004] uses SVM to detect cancer in pixel-level classifications segmenting regions of the cell potentially identifying cancer, [Amole and Osalusi 2018] uses both SVM and k-NN shallow classifiers to predict cancer in binary configurations achieving over 88% accuracy for both models. With this in mind, in this study we propose the use of SVM and EfficientNetB1 to detect cancer in pap smear screenings in both a binary and a 6-class classification, comparing the results of each approach.

3. Methodology

3.1. Image preprocessing

The images used for the Pap smear were grouped by class, being either one of (0) "Negative for intraepithelial lesion", (1) "ASC-US", (2) "ASC-H", (3) "LSIL", (4) "HSIL", or (5) "SCC". The total length of the initial dataset is of 5581 images separated unevenly between the aforementioned classes. The initial configuration of the images were as displayed in Table 1.

Table 1. Initial configuration of images in the dataset used

Class name	Class ID	Length
Negative for intraepithelial lesion	0	4244
ASC-US	1	84
ASC-H	2	391
LSIL	3	287
HSIL	4	493
SCC	5	82

All these images were cropped in a 100x100 configuration around the nucleus of the cells and separated each in its own folder with their respective identifiers in the original database.

3.2. Classifiers

The classifiers implemented had a combination of different approaches, totalling 4 combinations of tests. We classified pap smears using the Haralick Descriptors of entropy, homogeneity, and contrast, from the co-ocurrence matrices of $C_{1,1}$, $C_{2,2}$, $C_{4,4}$, $C_{8,8}$, $C_{16,16}$, and $C_{32,32}$, (a total of 18 features) for a shallow classifier; and we classified them using the matrices representing the pixels from the colored images, with 3 channels and 10,000 pixels (a total of 30,000 features) for a deep classifier. Both these approaches were applied in a binary context, where based on Table 1, we have class 0 and being negative, and all others as being positive, as well as in the full version, with all 6 classes. And finally, as for the classifiers itself, we used a shallow classifier, SVM, and a deep classifier, EfficientNetB1.

The training of all classifiers except the binary EfficientNetB1 classifiers included the augmented data from Table 1 in which the original dataset ended up 4 times larger. All original images were rotated clockwise in four angles: 0 (original image), 90, 180,

and 270. Since the original image cropping intentionally maintain the cell nucleus in the center, no aditional cropping or image edition is applied in the original dataset.

We compared all combination metrics in terms of accuracy, loss (for EfficientNetB1), and confusion, through the prediction's confusion matrices. The SVM was implemented with the hyperparameters combinations of kernels *rbf* and *poly*, regularization parameter (C) of 0.1, 1, and 10. For the kernel *poly*, we tested alongside these combinations the degrees of 1 (linear), 2, 3, 4, 5, 6, and 7; and the EfficientNetB1 was implemented with the hyperparameters combinations of batch size 4, 8, 16, 32, 64 and 128, learning rate of 0.001, 0.0001, 0.00005 and 0.00002 and 0.00001, 0, 1, 2, 3, and 4 hidden layers with 0, 4, 8, 16, 32, 64, 128, 256, and 512 neurons each with dropout of 50%, optimizers Adam and SGD (Stochastic Gradient Descent) all with early stopping and learning rate reducers.

4. Results and Discussion

The calculation of the 18 features for the entire dataset included calculating the 6 coocurrence matrices for each image, and the result is used to then calculate the entropy, homogeneity and contrast. This calculation took over 1 hour runnning in Google Colab. The training time for the best shallow model was of around 30 minutes, meanwhile the training of the deep classifier took well over 3 hours (without use of GPU acceleration) to complete.

Figure 1 was used as an example to demonstrate the results obtained by all four model implementations. The cell has the class 1 in both binary and multiclass classifications, with the binary SVM answering 1 (correctly indicating cancer); the multiclass SVM answering 1 (correctly indicating ASC-H); the binary EfficientNetB1 answering 1 (correctly indicating cancer); and the multiclass EfficientNetB1 answering 0 (incorrectly indicating no cancer). For this image in particular, 3 out of 4 models answered correctly.

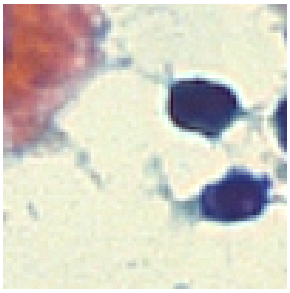

Figure 1. Example of image with ASC-H being classified by all four models

The best results for SVM were the parameters of kernel *poly*, regularization parameter of 10, and degree of 7, for both the binary and multi-class version, with 64% and 47% accuracy respectively. Meanwhile, for the EfficientNetB1 parameters, the best results had for the binary version with a batch size of 16, initial learning rate of 0.00001, 3 hidden layers with 8 neurons without dropout and optimizer Adam; and the same configurations for the multi-class version except the use of batch size of 128, 2 hidden layers with 512 neurons with 50% dropout. The normalized confusion matrix for the best combinations of both SVM and EfficientNetB1 are displayed in Figure 2.

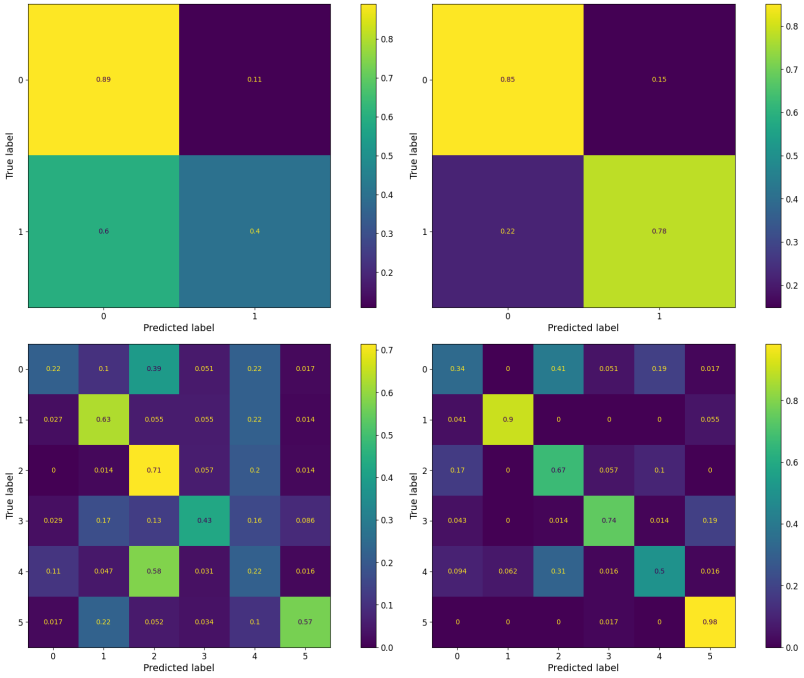

Figure 2. Confusion matrices of best models Binary SVM (top left) and EfficientNetB1 (top right), and multi-class SVM (bottom left) and EfficientNetB1 (bottom right)

As for the results of EfficientNetB1, we have a multi-class classifier that by itself is better than the binary SVM, with 65% accuracy, and a binary version with 82% accuracy. And by simply adapting the multi-class EfficientNetB1 to consider any prediction of classes 1, 2, 3, 4, or 5 as being 1 and treating the problem as binary with the multi-class classifier we are able to reach an average of 85% accuracy on an average of 10 runs in a random data sample. The training and validation accuracy and loss are displayed in Figure 3.

Overall, the results for the binary EfficientNetB1 were underwhelming, since the binary adaptation of the multi-class EfficientNetB1 was able to get a better average accuracy. The training of the binary EfficientNetB1 did not present very reliable results with the training metrics in Figure 3 not being maintained after loading the model (dropping from 97% to 82%).

Our results indicate that Deep Learning approaches with EfficientNetB1 are better suited for pap smears than the shallow classifiers with SVM. The multi-class version of our deep classifier reached a higher accuracy than the binary version of our shallow classifier as observed in Figure 2.

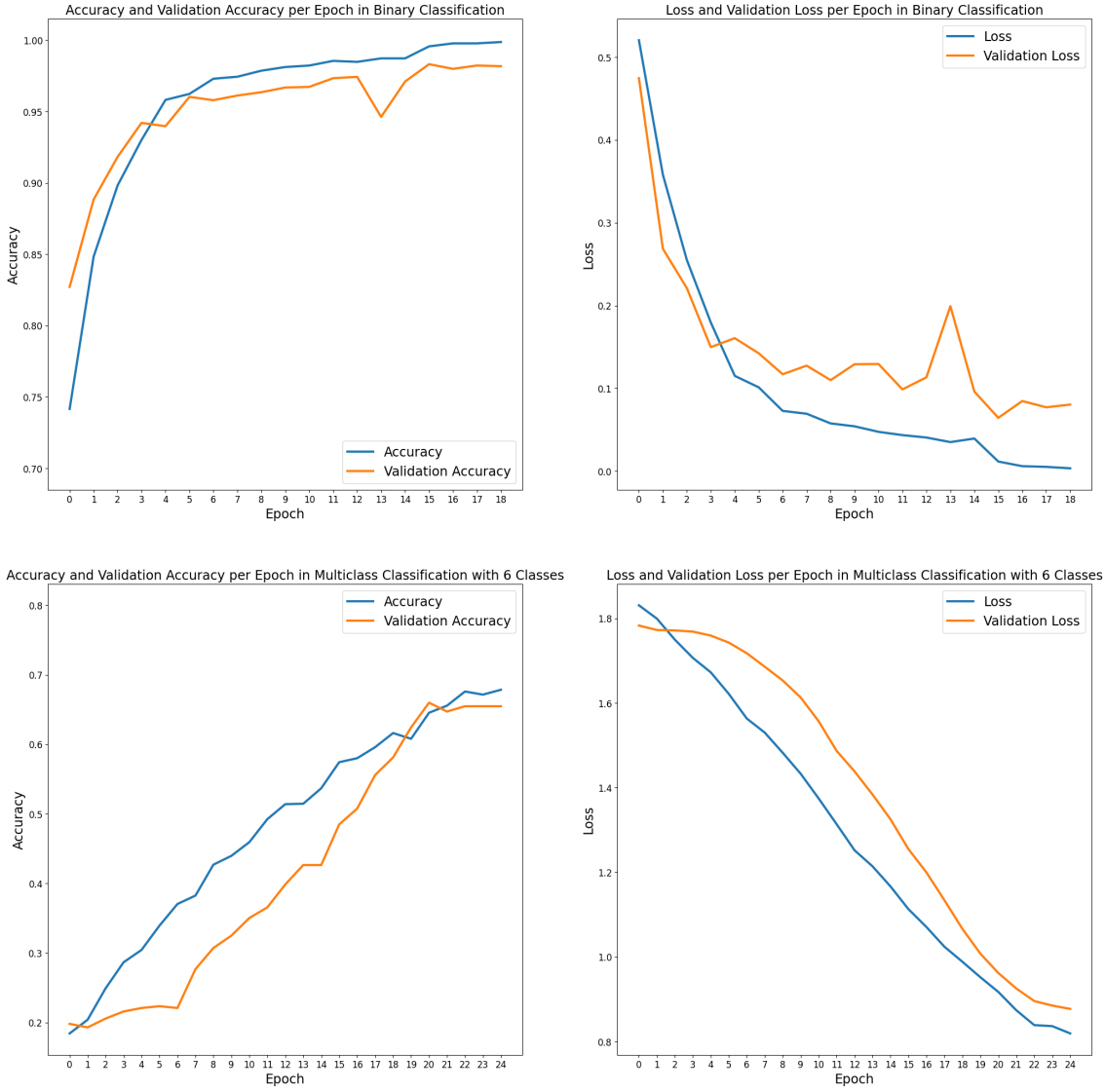

Figure 3. Training and validation metrics for binary and multi-class EfficientNetB1

5. Conclusion

We conclude that Deep Learning approaches are well suited for image classifications of pap smears, possibly reaching well over 85% accuracy in the binary configuration for cervical cancer identification. Meanwhile, the shallow classification with SVM reached at most 65% accuracy in the binary configuration. Future steps in this study include balance the dataset with a downsampler with probabilities estimator to avoid losing vital information in images removed because of random downsampling in an attempt of improving the final accuracy.

References

Amole, A. and Osalusi, B. S. (2018). Textural analysis of pap smears images for k-nn and svm based cervical cancer classification system. *Advances in Science, Technology and Engineering Systems Journal*.

Chen, X., Pu, X., Chen, Z., Li, L., Zhao, K.-N., Liu, H., and Zhu, H. (2023). Application of efficientnet-b0 and gru-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions. *Cancer Medicine*, 12(7):8690–8699.

Clarke, E. and Anderson, T. (1979). Does screening by "pap" smears help prevent cervical cancer?: A case-control study. *The Lancet*, 314(8132):1–4. Originally published as Volume 2, Issue 8132.

Koss, L. G. (1989). The Papanicolaou Test for Cervical Cancer Detection: A Triumph and a Tragedy. *JAMA*, 261(5):737–743.

Naucler, P., Ryd, W., Törnberg, S., Strand, A., Wadell, G., Elfgren, K., Rådberg, T., Strander, B., Johansson, B., Forslund, O., Hansson, B.-G., Rylander, E., and Dillner, J. (2007). Human papillomavirus and papanicolaou tests to screen for cervical cancer. *New England Journal of Medicine*, 357(16):1589–1597.

Rastogi, P., Khanna, K., and Singh, V. (2023). Classification of single-cell cervical pap smear images using efficientnet. *Expert Systems*, 40(10):e13418.

Rezende, M. T., Silva, R., Bernardo, F. d. O., Tobias, A. H., Oliveira, P. H., Machado, T. M., Costa, C. S., Medeiros, F. N., Ushizima, D. M., Carneiro, C. M., et al. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific data*, 8(1):151.

Zhang, J. and Liu, Y. (2004). Cervical cancer detection using svm based feature screening. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 873–880. Springer.