Original scientific paper

UDK 007.52:[591.483:681.32.06
DOI 10.7251/COMEN2202169H

COMPUTER AIDED DIAGNOSTIC SYSTEM FOR WHOLE SLIDE IMAGE OF LIQUID BASED CERVICAL CYTOLOGY SAMPLE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

Igor Hut[1*], Branislava Jeftić[1], Aleksandra Dragičević[1], Lidija Matija[1], Đuro Koruga[1,2]

[1] University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia
[2] TFT Nano Center, Belgrade, Serbia

[*] Corresponding author: ihut@mas.bg.ac.rs

Abstract: Cervical cancer screening with Papanicolaou test and liquid based cytology relies on the expertise of the pathologist. Liquid based cytology is proven to be more efficient than conventional Papanicolaou test when it comes to sample preparation and possibility of conducting several tests on the same sample. However, specificity and sensitivity of the test are in the range of the Papanicolaou test accuracy metrics, with false negative results still being the main drawback of these manually performed tests. Advances in technology and availability of digital data have enabled succesfull application of machine learning models in diagnostics. Images of cervical cells are now used as input to different deep learning models currently tested in studies concerning computer aided diagnostic systems. This study explores different architectures of convolutional neural network for cervical cancer detection based on Optomagnetic imaging spectroscopy and liquid based cytology samples. The proposed VGG16 based model achieved 93.3% sensitivity and 67.8% specificity in the binary classification problem. Results highlight the need for more balanced dataset in order for suggested deep model to achieve better performance.

Keywords: cervical cancer, liquid based cytology, convolutional neural network, optomagnetic imaging spectroscopy.

1. INTRODUCTION

Cervical cancer is the fourth most common cancer in women despite the success of the conventional diagnostic and screening tests made in last fifty years. The reason is contained in the fact that the practical implementation of the screening tests in low and middle income countries is often difficult due to lack of trained medical personnel and necessary infrastructure. Insufficient resources, absence of effective screening programs and adequate treatment have led to the occurrence of 84-90% worldwide deaths attributed to cervical cancer in these countries [1].

HPV test, Papanicolaou test and liquid based cytology (LBC) are primary screening tests used in developed countries. They are performed as a preventive measure to find abnormalities still not expressed in the form of cervical cancer symptoms. Liquid based cytology was primarily developed to overcome conventional cytology limitations. According to the reports of studies comparing conventional cytology to liquid based cytology, sample preparation by LBC systems offer several advantages over Papanicolaou smears out of which the removal of unwanted debris such as mucus, blood and other background artifacts is the most valuable, leading to higher rate of samples adequate for microscopic inspection. However, the same studies offer diverse conclusions when it comes to deciding whether LBC is more accurate than Papanicolaou

test [2-5]. Sensitivity of Papanicolaou test is relatively low and ranges from 55% to 80% for detection of cervical dysplasia and invasive cancer, compared to the 98% specificity of the test [6]. Compared to Papanicolaou test, liquid based cytology has similar sensitivity and specificity for the detection of cervical intraepithelial neoplasia grade 2 (CIN2+) and higher [7]. Important factors that lead to high number of misdiagnosis are errors occurring in the processes of sample collection, preparation and interpretation. High expertise of the cytopathologist is essential in order to overcome the subjectivity of the test. Another approach aimed to decrease the number of false negative results is integration of computer aided diagnostic systems in clinical settings. Main contribution of computer assisted screening for cervical smears would be reduction of pathologist's workload that would lead to minimizing the source of diagnostic error.

Papanicolaou test and LBC are offering the possibility of capturing cervical cell images which can then serve as a medium used in automated cancer detection systems. The number of studies investigating machine learning models for cancer detection is rapidly increasing, especially those concerning deep learning methods. Valuable diagnostic information can now be extracted from medical images, thanks to the technology advances and improved deep learning techniques available. With the help of these new technologies and automated classification models, the percentage of false negative results could be reduced. Still, most of the proposed deep learning models are based on the single cell cervical cytopathology images, not whole slides; they are not generalized for the classification involving cell overlapping and they are only tested on few available datasets [8, 9].

This study explores different architectures of convolutional neural network for cervical cancer detection using liquid based cervical cytology samples. Optomagnetic imaging spectroscopy was used for obtaining sample spectra and spectra were then subjected to deep model developed for classification of samples into one of the two defined categories. One category was made of negative cases, borderline (BL) and low grade squamous intraepithelial lesion (LSIL) cases and the other category consisted of high grade squamous intraepithelial lesions (HSIL) and cancer cases.

2. MATERIAL AND METHODS

2.1. Liquid Based Cytology samples

Cervical samples were collected from Southend University Hospital, UK and Oncquest Laboratories, New Delhi, India. A total of 535 cervical smears prepared according to liquid based cytology procedure were used in the study (Table 1). ThinPrep technology was used for LBC slide preparation: cervical cells were collected with cervical brush, and then rinsed into the PreservCyt transport medium. Preserved cells were then processed with semi-automated ThinPrep method, followed by subsequent staining and microscopic evaluation. All smears diagnosed as cancer or cervical intraepithelial neoplasia were confirmed with histopathology.

Table 1. *Distribution of liquid based cervical cytology smears and number of considered samples from each category*

Diagnosis	Number of cervical smears	Classification category
Normal	327	low risk
Borderline	32	
LSIL	97	
HSIL	64	high risk
Cancer	15	

2.2. Optomagnetic Imaging Spectroscopy

Optomagnetic imaging spectroscopy was used to obtain optomagnetic spectra of the samples. The procedure of spectra collection includes slide screening with OMIS device and image processing with OMIS convolutional algorithm [10]. OMIS device uses Canon camera IXUS 133, 16 MP, image stabilization, with double/active lens for capturing digital images of the cervical samples under the white diffuse light and polarized white light. Digital images are processed in Matlab® 2013a (MathWorks, USA).

2.3. Deep neural network architecture

The general type of deep neural network used for this study is a Convolutional Neural Network (CNN/ConvNet), which is the foundation of modern computer vision. More specific, the utilized CNN is VGG16. It is a ConvNet model developed by Karen Simonyan and Andrew Zisserman at the University

Table 2. *VGG configurations [11].*

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

of Oxford. The idea of the model was proposed in 2013, but the actual model was submitted during the ILSVRC ImageNet Challenge in 2014 [11]. The name VGG comes after the department of Visual Geometry Group in the University of Oxford that authors belonged to. Authors initially proposed several VGG configurations, as shown in the Table 2.

A stack of multiple (usually 1, 2, or 3) convolution layers of filter size 3 x 3, with stride of one, and padding of one, followed by a max-pooling layer of size 2 x 2, is the basic building block for all these configurations. Different configurations of this basic stack are repeated for various network configurations to achieve different network depths. The number of layers containing weight parameters determines the number associated with the specific configuration. The convolution stacks are followed by three fully connected layers, two with the size of 4,096 and the last one with the size of 1,000 nodes. The last layer is the output layer with SoftMax activation. The size of 1,000 refers to the total number of possible classes in the ImageNet dataset. Hence, VGG16 refers to the configuration "D" in table. The configuration "C" also has 16 weight layers, but it uses a 1 x 1 filter as the last convolution layer in stacks 3, 4, and 5. This layer was used to increase the non-linearity of the decision functions without affecting the receptive field of the layer. The configuration "D" will be referred as VGG16 in the further discussion. Conceptual graphical representation of VGG16 model architecture and layer scheme can be seen in Fig. 1 and Fig. 2. The input to any VGG16 configuration is a fixed size 224x224 RGB, i.e the input needs to be a tensor of 224x224x3 size. This input is passed through the first stack of 2 convolution layers with a very small receptive field, i.e., 3 × 3 (which is the smallest size to capture the notion of left/right, up/down, and center), followed by

ReLU activations. Each of these two layers contains 64 filters. The convolution stride and the padding are fixed at 1 pixel. Such configuration preserves the spatial resolution, hence dimensions of the output activation map are the same as the input image's. Further, the activation maps are passed through spatial max pooling over a 2 x 2-pixel window, with a stride of 2 pixels, which halves the size of the activations. This brings the size of the activations at the end of the first convolutional stack to 112 x 112 x 64.

The second convolutional stack has 128 filters, hence the size of its output becomes 56x56x128. The third stack consists of 3 convolutional layers with 256 filters and a max pool layer, which brings the output of this stack to 28x28x256. This is followed by two stacks of three convolutional layers, with each containing 512 filters. Size of the output for both stacks will be 7x7x512.

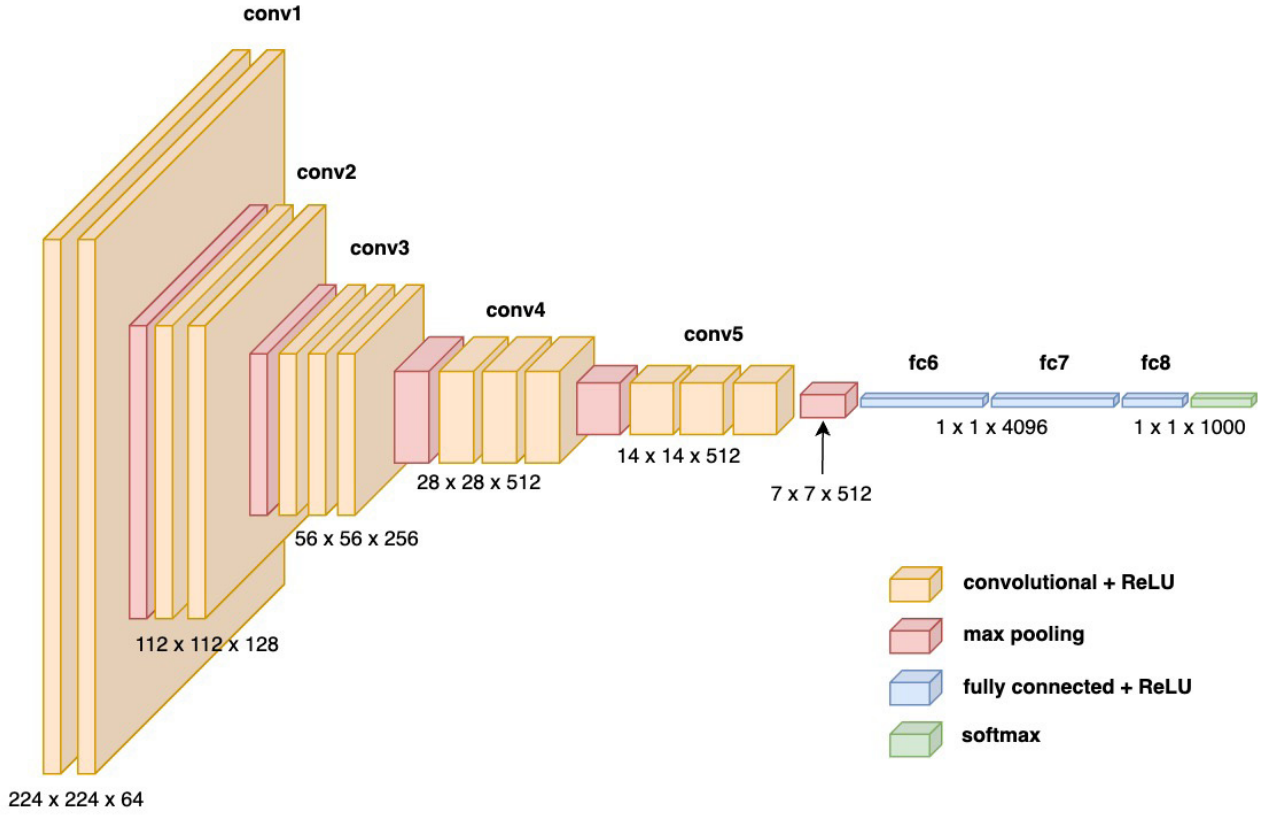

Figure 1. *VGG16 ConvNet architecture representation.*

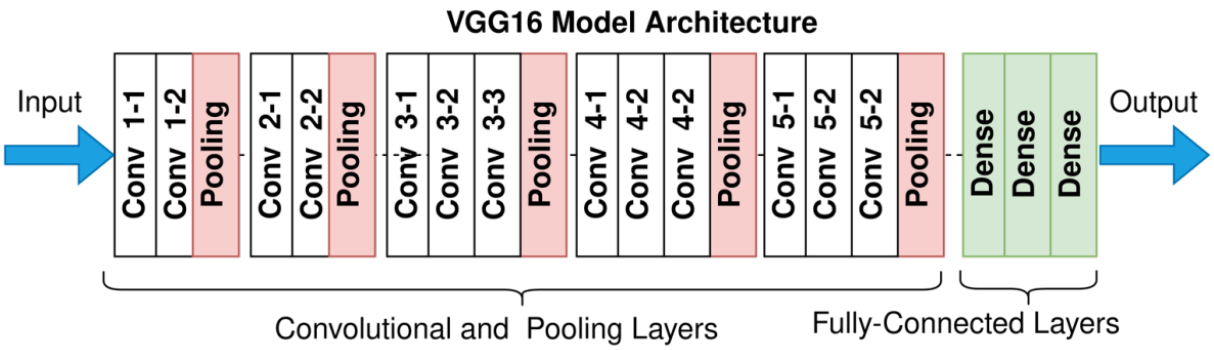

Figure 2. *VGG16 ConvNet layers scheme.*

The network ends with three fully connected layers, with a flattening layer in-between. The first two layers contain 4096 neurons each, whilst the last fully connected layer has 1000 neurons corresponding to the 1,000 possible classes for the ImageNet dataset. This last layer serves as the output layer. Output goes into the softmax activation layer, used for categorical classification.

To build an adequate model for our study, the transfer learning approach has been utilized. Transfer learning approach means that a model trained on one set of data is used as a starting point for modeling a new set of data. More precisely pre-trained network is used to extract features from the new data set, that will be used to train a new, relatively simple, neural network for the end task of classification. In other words, the convolutional base is reused while adding a new densely connected classifier structure. Representations learned by convolutional base, on large and diverse datasets, are generic enough to be reusable, i.e., the feature maps of a CNN are presence maps of generic concepts over a picture, which are likely to be useful regardless of the specific computer vision problem.

For this study two different fully connected architectures have been utilized to complete the network, by adding them on top of the VGG16 convolutional base.

The first one is comprised of following layers: 256 neurons with ReLU activation; Dropout layer with 20% dropout rate to prevent overfitting; batch normalization, i.e. Batch Norm, layer which normalizes the output of the dropout layer speeding up the learning and, also, decreasing the risk of overfitting; layer with 4096 neurons with ReLU activation followed by one more Batch Norm layer. The last layer ends with softmax function which proved to be the best "standard" option for final classification. The softmax activation function transforms the outputs of the neural network into a vector of probabilities, essentially giving a probability distribution over the input classes. The final CNN containing this dense architecture is denoted as "VGG16 4096" (CNN model 1) in the following discussion.

The second dense architecture used in this study contains the following layers: 128, 64 and then again 128 neurons all with ReLU activation, each followed by dropout layers with 20% dropout rate and Batch Norm layer. Final layer, again, is completed with the softmax activation function at the end. This configuration is referred to as "VGG16 2x128" (CNN model 2) in the rest of the analysis.

3. RESULTS AND DISCUSSION

Two VGG16 architectures were tested on Optomagnetic spectra of the LBC cervical samples: VGG16 4096 (CNN model 1) and VGG16 2x128 (CNN model 2). Fine tuning process of VGG16 4096 model during 20 training epochs is shown on Figure 3.

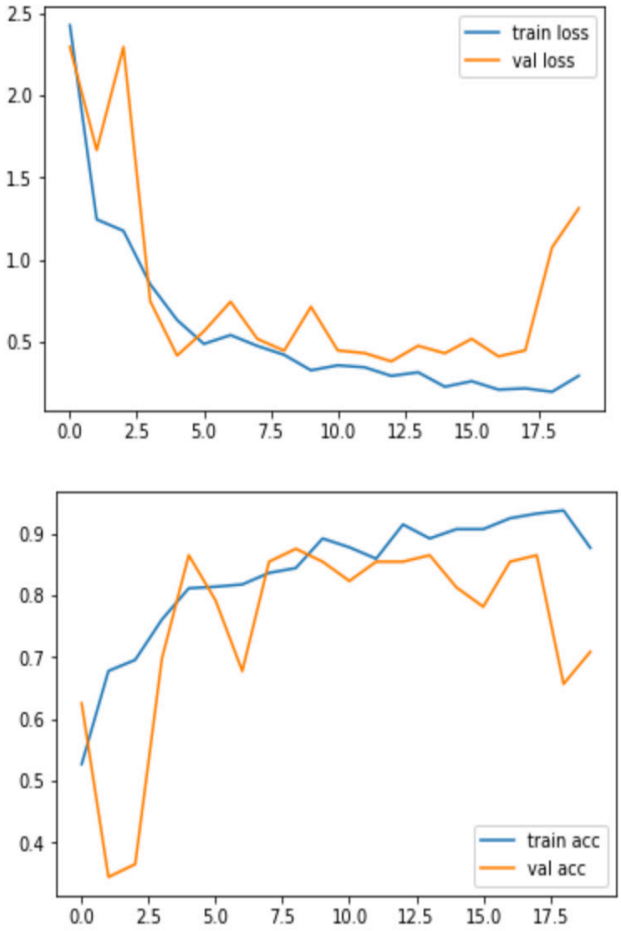

Figure 3. *VGG16 4096 model: training and validation loss and accuracy*

VGG16 4096 arhitecture achieved high sensitivity (93.3%), however high number of false positives has contributed to lower specificity of 67.8% and overall accuracy of 70.8% (Figure 4).

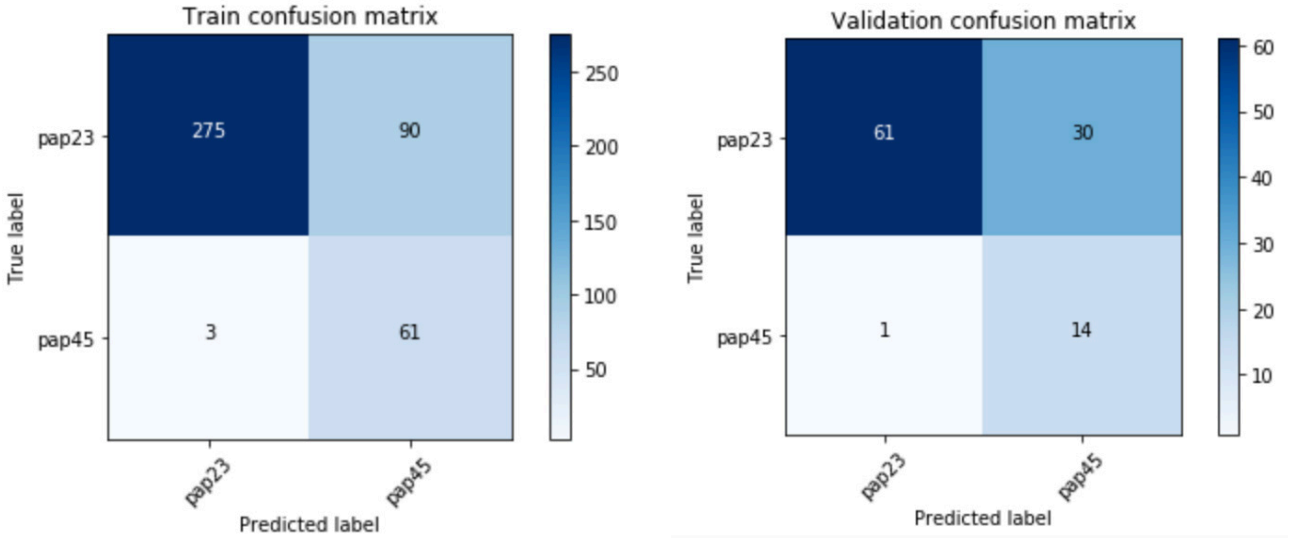

Figure 4. *VGG16 4096 model: confusion matrix*

Among 91 low risk cases, 30 cases were labeled by model 1 as high risk (32.97%), and among 15 high risk cases, 1 was incorrectly labeled as low risk (6.67%).

Fine tuning process of VGG16 2x128 model during 20 training epochs is shown on Figure 5.

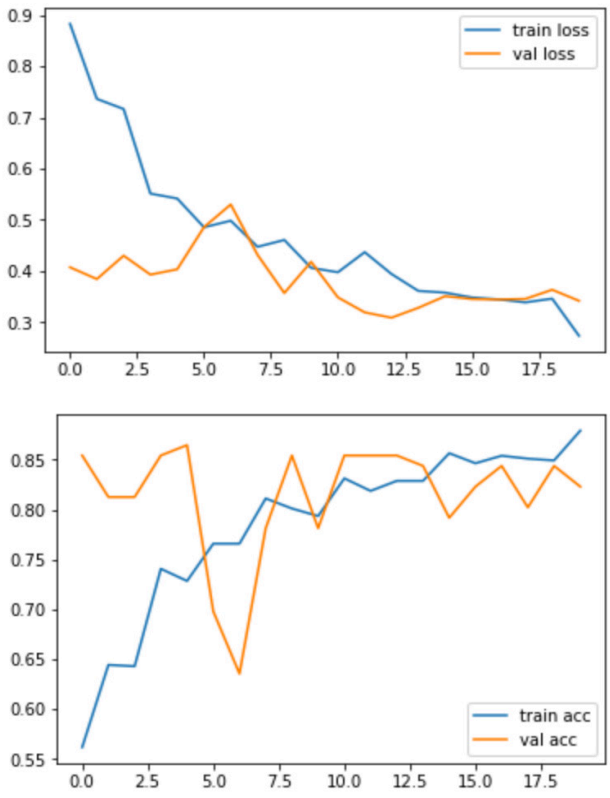

Figure 5. *VGG16 2x128 model: training and validation loss and accuracy*

According to the results given in confusion matrix (Figure 6), VGG16 2x128 model didn't perform well on the given dataset in terms of sensitivity, since it achieved sensitivity of only 26.6%. Among 91 low risk cases, 8 cases were labeled by model 1 as high risk (8.8%), and among 15 high risk cases, 11 were incorrectly labeled as low risk (73.33%).

The evident problem is small number of data, as well as uneven distribution of data in two defined classes. Deep neural network is tested on 535 images, and high risk class is approximately six times smaller than low risk class. In the study that tested several classical machine learning algorithms on the set of 700 LBC smear optomagnetic spectra has achieved highest accuracy with Random Forest model, 79% sensitivity and 83% specificity. Binary classification was used to separate normal from abnormal cases and distribution of the cases was even across the classes (there were 354 negative cases in the first "normal" class, and 346 cases in second, "abnormal" class that consisted of moderate dysplasia, HSIL, severe dysplasia and cancer cases) [12]. However, in order to classify samples with machine learning algorithm, manual extraction of the features is needed to be done prior to classification task. Deep learning approach is based on automatic extraction of significant features, and considers information within the dataset that may be overlooked by human experts. Zhang et al proposed deep neural network for cervical cancer classification and tested it on Papanicolaou smears and LBC samples. Even though they report high classification perfor-

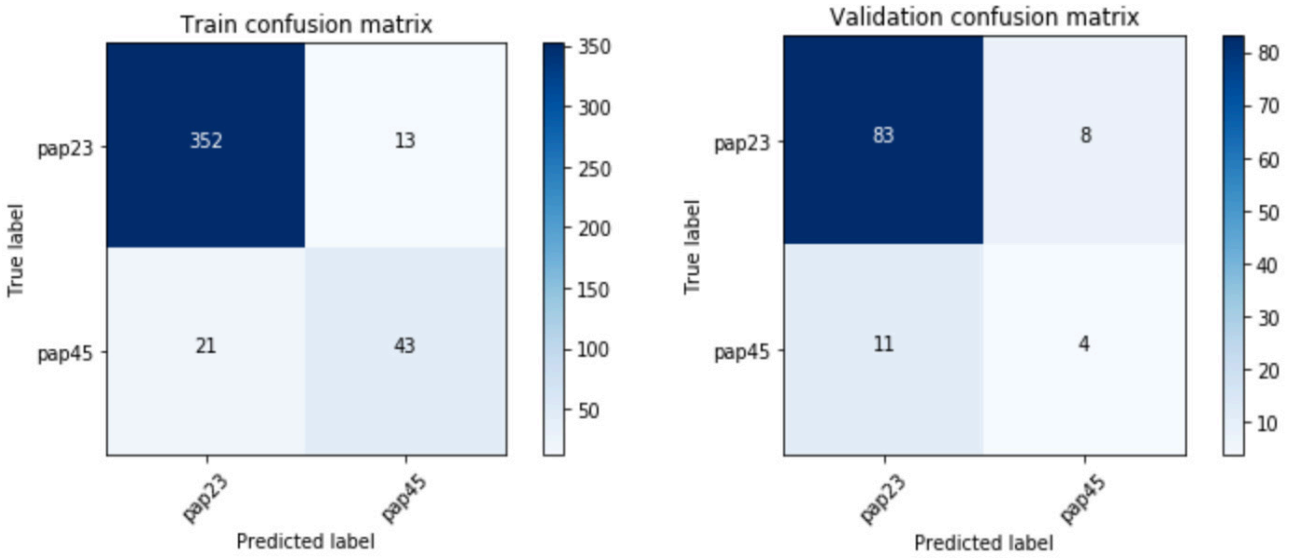

Figure 6. *VGG16 2x128 model: confusion matrix*

mance of their method (98.2% sensitivity, 98.3% specificity and 98.3% accuracy), limitation of the study is that the system they propose isn't optimal for clinical implementation since it requires specifically crafted images. Moreover, dataset used in this study comprises only of images of individual cells, thus avoiding real case scenarios (ex. overlapping cells) [13]. Sanyal et al tested two CNN models on LBC smears, and reported sensitivity of 95.63% and specificity of 79.85%. Although the dataset used in this study consisted of 2816 images, more diverse dataset is needed in order for this system to be useful as diagnostic tool [14]. Despite the high accuracy percentages reported in current algorithms based on deep learning, the lack of solutions for whole-slide analysis, different approaches to staining and imaging, as well as validation on clinical datasets, still remains the challenging problem. Some studies offer solution for the analysis of whole slides with the robust deep learning model generalized for various types of staining and imaging, still the question of the class definition remains unresolved; some focuses on classification into positive/negative class, others into more subclasses [15-17]. Our method is based on utilizing cervical cell magnetic properties captured in the form of Optomagnetic spectrum, whereby all cervical cells in the smear are considered, which is equivalent to the whole slide inspection. The automatized process of obtaining Optomagnetic spectra is finished within 2 minutes, and no additional preprocessing is needed prior to classification. High sensitivity of the proposed VGG16 4096 network architecture and the ability to properly classify high risk cases indicate its potential as a screening tool that would serve for identifying cases in need for further testing.

4. CONCLUSION

Conventional screening tests, such as Papanicolaou test and liquid based cytology have greatly contributed to the decrease of cervical cancer incidence and mortality rate, still there is high number of false negative results related to them. Moreover, such screening tests are usually difficult to implement in low resource settings, where most of the death cases attributed to cervical cancer occur. Digital images of cervical cytology smears are proving to be convinient tool for automated deep learning detection of cervical cancer. Despite a large number of recent studies reporting high accuracies of deep learning models used for cervical cancer detection, only few of them are resolving some of the key problems emerging within this topic. The most important are dealing with the fact that most of the models are trained and tested on specific datasets, usually 2-3 available, so the question remaines if the classification results would be similar if models were tested on clinic specific images, especially if we agnowledge the fact that clinics use different staining procedures, different imaging systems, magnifications, etc. Furthermore, most of the proposed solutions are based on the classification of single cell images. Our deep learning model considers whole

smear through optomagnetic spectra of the cervical samples, and classifies samples into low and high risk class with the sensitivity of 93.3% and specificity of 67.8%. Low specificity is due to small and highly imbalanced dataset, thus in further investigations larger and more balanced dataset will be required to improve specificity and sensitivity of the model.

5. ACKNOWLEDGEMENT

This research has been partialy funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, through the Project No. 451-03-68/2022-14/200105 from 04-Feb-2022. Special thanks go to Dr Khalil Razvi, Southend University Hospital, UK, and Dr. Ravi Gaur, Oncquest Laboratories (New Delhi), India for their professional support, and to Aleksandar Obradovic, Tumour Trace, UK/Serbia, whose financial support enabled field research and data acquisition in UK and India.

6. REFERENCES

[1] R. Hull, et al., *Cervical cancer in low and middle-income countries*, Oncology letters, 20-3, (2020) 2058-2074.

[2] E. Davey, A. Barratt, L. Irwig, S.F. Chan, P. Macaskill, P. Mannes, A.M. Saville, *Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review*, The Lancet, 367-9505 (2006) 122-132.

[3] M. Arbyn, C. Bergeron, P. Klinkhamer, P. Martin-Hirsch, A.G. Siebers, J. Bulten, *Liquid compared with conventional cervical cytology: a systematic review and meta-analysis*, Obstetrics & Gynecology, 111-1 (2008) 167-177.

[4] A.G. Siebers, et al., *Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial*, Jama, 302-16 (2009) 1757-1764.

[5] S. Pankaj, A. Kumari, S. Kumari, V. Choudhary, J. Kumari, A. Kumari, S. Nazneen, *Evaluation of sensitivity and specificity of pap smear, LBC and HPV in screening of cervical cancer*, Indian Journal of Gynecologic Oncology, 16-3 (2018) 1-5.

[6] S.L. Bedell, L.S. Goldstein, A.R. Goldstein, A.T. Goldstein, *Cervical cancer screening: past, present, and future,* Sexual medicine reviews, 8-1 (2020) 28-37.

[7] V. Bouvard, et al., *The IARC perspective on cervical cancer screening, New England Journal of Medicine, 385*-20 (2021) 1908-1918.

[8] A. Khamparia, et al., *Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning,* The Journal of Supercomputing, 76-11 (2020) 8590-8608.

[9] H. Chen, et al., *CytoBrain: cervical cancer screening system based on deep learning technology*, Journal of Computer Science and Technology 36-2 (2021) 347-360.

[10] Dj. Koruga, A. Tomic, MYSKIN Inc, 2009. System and method for analysis of light-matter interaction based on spectral convolution. U.S. Patent Application 12/350,164.

[11] K. Simonyan, A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. *arXiv preprint arXiv:1409.1556* (2014).

[12] I. Hut, B. Jeftic, L. Matija, Z. Cojbasic, Dj. Koruga, *Machine learning classification of cervical tissue liquid based cytology smear images by optomagnetic imaging spectroscopy*, Tehnički vjesnik, 26-6 (2019) 1694-1699.

[13] L. Zhang, L. Lu, I. Nogues, R.M. Summers, S. Liu, J. Yao, *DeepPap: deep convolutional networks for cervical cell classification*, IEEE journal of biomedical and health informatics, 21-6 (2017) 1633-1643.

[14] P. Sanyal, S. Barui, P. Deb, H.C. Sharma, *Performance of a convolutional neural network in screening liquid based cervical cytology smears,* Journal of cytology, 36-3 (2019) 146.

[15] H. Lin, H. Chen, X. Wang, Q. Wang, L. Wang, P.A. Heng, *Dual-path network with synergistic grouping loss and evidence driven risk stratification for whole slide cervical image analysis*, Medical Image Analysis, 69 (2021) 101955.

[16] S. Cheng, et al., *Robust whole slide image analysis for cervical cancer screening using deep learning,* Nature communications, 12-1 (2021) 1-10.

[17] F. Kanavati, N. Hirose, T. Ishii, A. Fukuda, S. Ichihara, M. Tsuneki, *A Deep Learning Model for Cervical Cancer Screening on Liquid-Based Cytology Specimens in Whole Slide Images*, Cancers, 14-5 (2022) 1159.

КОМПЈУТЕРСКИ ДИЈАГНОСТИЧКИ СИСТЕМ ЗА КЛАСИФИКАЦИЈУ СЛИКА ЦЕРВИКАЛНИХ ЦИТОЛОШКИХ РАЗМАЗА НА ТЕЧНОЈ БАЗИ ПРИМЕНОМ КОНВОЛУЦИОНИХ НЕУРОНСКИХ МРЕЖА

Сажетак: Скрининг карцинома грлића материце Папаниколау тестом и цитологијом на течној бази зависи од експертизе патолога. Показано је да цитологија на течној бази обезбеђује већи проценат препарата задовољавајућих за анализу у поређењу са Папаниколау тестом и да пружа могућност извођења додатних тестова на истом узорку. Међутим, специфичност и сензитивност цитологије на течној бази налазе се у опсегу мера тачности Папаниколау теста и лажно негативни резултати и даље представљају слабу страну ових мануалних дијагностичких тестова. Технолошки напредак и доступност дигиталних података отворили су пут успешној примени модела машинског учења у дијагностици. Слике цервикалних ћелија данас се користе као улаз у различите моделе дубоког учења тестиране у студијама које се баве компјутерским дијагностичким системима. Ова студија испитује различите архитектуре конволуционих неуронских мрежа за потребе детекције канцера грлића материце на бази Оптомагнетене имиџинг спектроскопије и цервикалних цитолошких узорака на течној бази. Предложени модел на бази VGG16 архитектуре постигао је сензитивност од 93.3% и специфичност од 67.8% у бинарном класификационом проблему. Резултати наглашавају потребу за скупом података који је уравнотежен како би се постигле боље перформансе предложеног дубоког модела.

Кључне ријечи: канцер грлића материце, цитологија на течној бази, конволуционе неуронске мреже, оптомагнетна имиџинг спектроскопија.

Paper received: 24 August 2022 **Paper accepted: 07 October 2022**