HISTOPATHOLOGICAL H&E-STAINED IMAGE ANALYSIS BASED ON AI

Jelena Musulin[1], Daniel Štifanić[1], Ana Zulijani[2], Sandi Baressi Šegota[1], Nikola Anđelić[1], Ivan Lorencin[1], Matko Glučina[3], Zlatan Car[1]

[1] Faculty of Engineering,
The University of Rijeka, Vukovarska 58, 51000 Rijeka
e-mail: jmusulin@riteh.hr, dstifanic@riteh.hr, sbaressisegota@riteh.hr, ilorencin@riteh.hr, nandelic@riteh.hr, car@riteh.hr
[2] Department of Oral Surgery,
Clinical Hospital Center of Rijeka, Krešimirova Ul. 40, 51000 Rijeka
e-mail: ana.zulijani@gmail.com
[3] University of Rijeka,
Trg braće Mažuranića 10, 51000 Rijeka
e-mail: matko.glucina@uniri.hr

Abstract:

Over the past decade, improvements in image analysis methods and substantial advancements in processing power have allowed the development of powerful computer-aided analytical approaches to medical data. Tissue histology slides can now be scanned and preserved in digital form, thanks to the recent introduction of entire slide digital scanners. In such form, they can serve as input data for Artificial Intelligence (AI) algorithms that can speed up standard procedures for histology analysis with high accuracy and precision. The aim of this research was to create an automated system based on AI for histopathological image analysis. The first step was normalizing H&E-stain images, then using them as input to the convolutional neural network. Such an approach proved to be successful in analyzing histopathological images.

Key words: artificial intelligence, convolutional neural network, histopathological analysis, oral squamous cell carcinoma

1. Introduction

Image data is extremely important in healthcare. Lately, the massive accumulation of digital images has increased the demand for their analysis, such as computer-aided diagnosis using Artificial Intelligence algorithms. Medical image analysis is one of the areas where histological tissue patterns are combined with computer-aided image analysis to improve detection and classification of disease [1]. There is also the possibility of automating and speeding up processes that take a long time to complete manually. AI-based models may learn to recognize specific traits in these images, making the diagnostic procedure much faster and more accurate.

The dataset used in this research consists of histopathological images of the oral squamous cell carcinoma region (OSCC), which contains abnormalities. Oral squamous cell carcinoma is the most common histological neoplasm of head and neck cancers, and while it is located in an easily visible area and can be detected early, this does not always eventuate [2]. Despite advances in therapeutic approaches, the morbidity and mortality rates from OSCC have not improved significantly over the last 30 years. The 5-year survival rate for patients with OSCC ranges between 40% and 50% [3]. The most prevalent reasons why OSCC is detected in advanced stages include an incorrect initial diagnosis and the ignorance from the patient or from the attending physician.

Clinical examination, conventional oral examination (COE), and histological evaluation following biopsy are procedures for detecting oral cancer. These procedures can detect cancer in the stage of established lesions with significant malignant changes. However, the subjective component of the examination, respectively inter- and intra-observer variability, is the fundamental difficulty in employing histopathological examination for tumor differentiation. Moreover, from the pathologist's point of view, providing exact histological identification in the context of multi-class grading is crucial. For this reason, combination of AI-based approaches with clinical prospective could reduce inter- and intra-observer variability as well as assist pathologists in terms of reducing the load of manuals inspection in shorter time [2].

2. Materials and Methods

For this research 127 histopathological H&E-stained images with 768 x 768-pixel size have been used to create the dataset. Haematoxylin and eosin (H&E) staining is one type of tissue staining that is of particular interest to pathologists. This is because the H stain highlights nuclei in blue against a pink cytoplasmic background (and other issue regions). This allows a pathologist to quickly identify and examine tissue, which is a labor-intensive operation. The OSCC samples were retrieved from the archives of the Clinical Department of Pathology and Cytology, Clinical Hospital Center in Rijeka. Two unbiased pathologists analyzed sample slides and classified them according to the 4th edition of the World Health Organization (WHO) classification of Head and Neck malignancies and the 8th edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual.

Since fields like medical image analysis rarely have access to a large number of samples, whilst AI models rely on a large number of samples to achieve good performance and avoid overfitting, it is required to use augmentation techniques to significantly improve the amount and quality of the data [4]. Geometrical transformations used for the augmentation procedure are horizontal flip, horizontal flip combined with 90 degrees anticlockwise rotation, vertical flip, and vertical flip combined with 90 degrees anticlockwise rotation, 90 degrees anticlockwise rotation, 180 degrees anticlockwise rotation and 270 degrees anticlockwise rotation. The augmentation process is used only for the development of training samples, as newly generated data are variants of the original data. Testing samples are not augmented.

Convolutional neural networks (CNN) have emerged as the most prominent strain of neural networks in research in recent years [5]. They have revolutionized computer vision, achieving cutting-edge results in many fundamental tasks while also making significant advances in natural language processing, reinforcement learning, and many other areas. In this research, for classification purposes, we are using one of the CNN models called VGG16. It is proposed by Simonyan & Zisserman, consists of 19 weight layers, and has been evaluated for large-scale image classification [6].

3. Results and Discussion

An automated H&E-stain histopathological analysis could assist the pathologist in discovering new informative features and in analyzing the tumor microenvironment. In order to perform automated image analysis, H&E-stained images need to be normalized. This is due to the large color variations in images caused by sample preparation and imaging settings. In our research, we used Macenko approach [7] where the Singular Value Decomposition (SVD) geodesic method is used for obtaining stain vectors. The first step is to convert the RGB color vector to their corresponding optical density (OD) values then remove data with OD intensity less than β. Threshold value of $\beta = 0.15$ was found to provide the most robust results while removing as little data as possible. The next step is to calculate singular value decomposition (SVD) on the OD tuples then create a plane from the SVD directions corresponding to the two largest singular values. After projecting data onto the plane and normalizing to the unit length we calculate the angle of each point with regard to the first SVD direction. The final step is to convert extreme values back to OD space. Figure 1. shows an image before and after normalization.

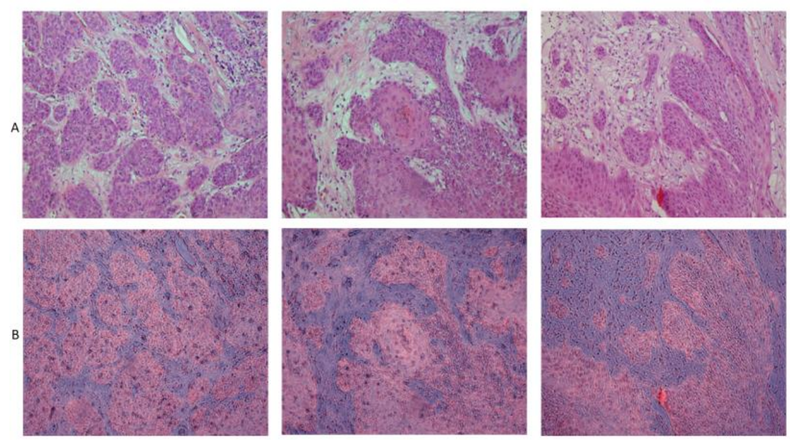

Fig. 1. Visual representation of A) H&E-stained images and B) normalized H&E-stained images

Stratified 5-fold cross-validation is used to estimate the performance of AI-based model while Area Under the ROC Curve (AUC) is used as evaluation metric.

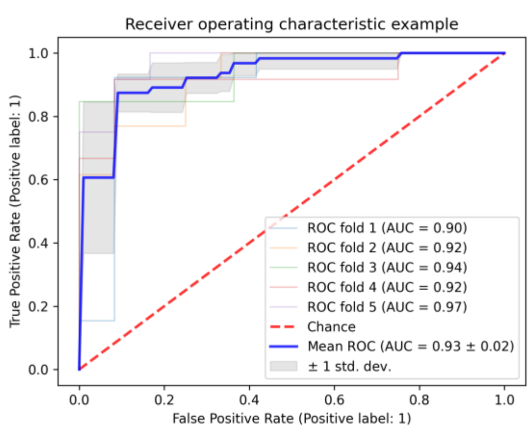

Fig. 2. Receiver operating characteristic (ROC) with cross validation

After image normalization VGG16 resulted in the highest classification value of 0.93 (σ ± 0.02) AUC. Figure 2. represents the ROC metric to evaluate classifier output quality using cross-validation. Shown are ROC curves for each of the 5-folds cross-validation and the overall average ROC curve (blue), based on VGG16 predictions. The best results were achieved when two additional layers were added at the end of the base VGG16 architecture.

4. Conclusions

Obtained results reveal that the application of AI-based algorithms along with preprocessing methods, such as image normalization for image analysis, has great potential in the diagnosis of OSCC. Integration of preprocessing method along with convolutional neural network resulted with 0.935 (σ ± 0.045) AUC. However, data availability was a limitation of the research so the future work should use a dataset with more histopathology images to create a more robust system. The presented approach is the first step in automating histopathological image analysis, therefore, in future work we plan to integrate more preprocessing methods with other AI classification algorithms.

Acknowledgments

This research has been (partly) supported by the CEEPUS network CIII-HR-0108, European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS), project CEKOM under the grant KK.01.2.2.03.0004, Erasmus+ project WICT under the grant 2021-1-HR01-KA220-HED-000031177 and University of Rijeka scientific grant uniri-tehnic-18-275-1447.

References

[1] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. IEEE reviews in biomedical engineering, 2, 147-171.

[2] Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers, 13(8), 1784.

[3] Zanoni, D. K., Montero, P. H., Migliacci, J. C., Shah, J. P., Wong, R. J., Ganly, I., & Patel, S. G. (2019). Survival outcomes after treatment of cancer of the oral cavity (1985–2015). Oral oncology, 90, 115-121.

[4] Musulin, J., Štifanić, D., Zulijani, A., Šegota, S. B., Lorencin, I., Anđelić, N., & Car, Z. (2021, October). Automated Grading of Oral Squamous Cell Carcinoma into Multiple Classes Using Deep Learning Methods. In 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 1-6). IEEE.

[5] Štifanić, D., Musulin, J., Car, Z., & Čep, R. (2020). Use of Convolutional Neural Network for Fish Species Classification. Pomorski zbornik, 59(1), 131-142.

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

[7] Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., ... & Thomas, N. E. (2009, June). A method for normalizing histology slides for quantitative analysis. In 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 1107-1110). IEEE.