УДК 004.04:[57+61]

ON WHOLE-SLIDE IMAGERY AND COMPUTATIONAL PATHOLOGY IN MEDICAL DIAGNOSIS

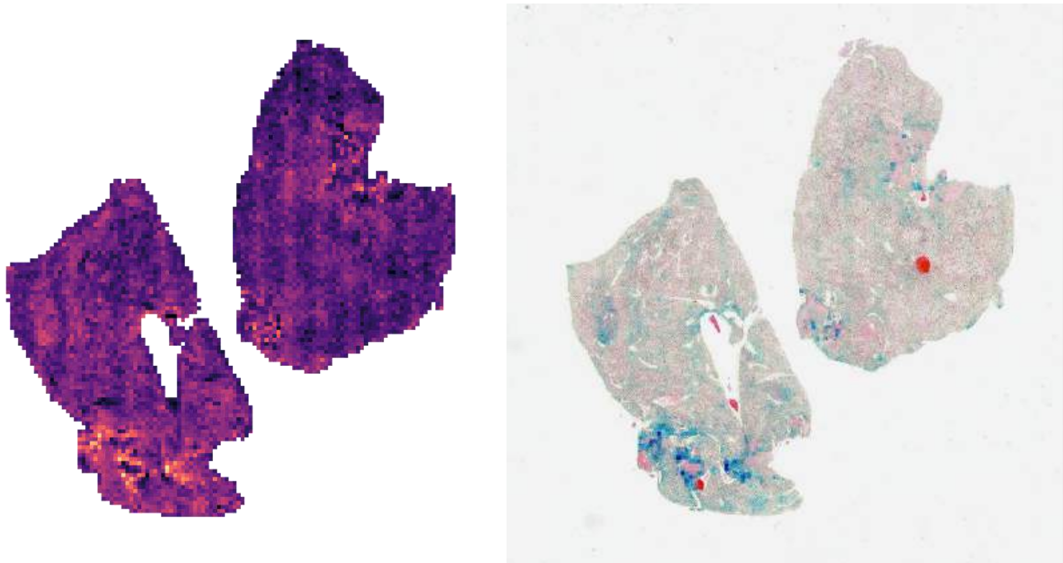
Канцемал А.О.
Науковий керівник – д.т.н., проф. Перова І.Г.
Харківський національний університет радіоелектроніки, каф. СТ
м. Харків, Україна
тел. +38(098) 465-03-25, email: anton.kantsemal@nure.ua.

This work is devoted to exploring the phenomena of whole-slide imagery and computational pathology in medical diagnosis. In addition, we assess CLAM – one of the methods of computational pathology, aimed to process whole-slide tissue scans – as a way of addressing the important whole-slide imagery processing difficulties.

> *Whole-slide digital imaging is the process by which digital images of entire histologic slides are created by high-resolution scanning of the slide. [1]*

In other words, a whole-slide image is a very high-resolution digital scan of a histology sample. The advancements in scanning technology in recent years allow us to get slides at the gigapixel scale. With that, in order for the pathologist to make a correct decision, they have to look through every pixel rigorously. On the other hand, modern methods of computational analysis enable us to process this much medical data in no time. This process is called computational pathology.

> *Computational pathology (CPATH) – a branch of pathology that involves computational analysis of a broad array of methods to analyze patient specimens for the study of disease. [2]*



Picture 1. Left: direct visualization of the attention scores. Each pixel represents a 256px patch. Right: attention scores blended onto the WSI. The colored regions represent the most attended parts of the sample.

Modern computational pathology is able to provide extensive support for pathological analysis. For example, features may be extracted to compactly represent the whole slide. Also, the said features may be used in classical ML algorithms to correlate whole-slide images to the patient outcome. In addition to that, CPATH techniques may lead to discovery of previously unrecognized morphological characteristics with clinical relevance that have not been used in visual assessment by pathologists, either because these features had not previously been discovered or because they are beyond human visual perception.

An example of CPATH – Clustering-constrained attention multiple instance learning, or CLAM for short – a work of Lu et al. [3], and is a data-efficient approach to working with weakly-labeled whole-slide imagery.

We have trained the CLAM model on a subset of publicly-available Clear Cell Renal Cell Carcinoma (CPTAC-CCRCC) dataset [4]. As we can see on the picture 1, the CLAM's attention network is quite interpretable. It outputs "importance scores" for each patch, so those can be visualized as a heatmap. If you take a closer look at the WSI on the right, you may see the affected cells colored blue, meaning the network decided them to affect the prediction the most. This approach addresses important WSI processing difficulties. For example, to handle the large size of the slide, we segment the actual tissue from the scan background, trimming unnecessary information. Then, we split the remaining ROI into tiles, that are treated equally, to fit them into the GPU. Also, another difficulty is labeling. It takes hours for a pathologist to look through the slide and diagnose, let alone label the cause of it, i.e., the affected cells. On the contrary, CLAM is designed to work with this kind of weakly-labeled data, where only the slide-level labels are known (i.e., diagnosis.) CLAM itself determines which regions of the tissue affect the predicted diagnosis.

Список використаних джерел:

1. Shaimaa Al-Janabi, André Huisman, Paul J Van Diest. Digital pathology: current status and future perspectives.

2. Esther Abels, Liron Pantanowitz, Famke Aeffner, Mark D Zarella, Jeroen van der Laak, Marilyn M Bui, Venkata NP Vemuri, Anil V Parwani, Jeff Gibbs, Emmanuel Agosto-Arroyo, Andrew H Beck, Cleopatra Kozlowski. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association.

3. Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, Faisal Mahmood. Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images.

4. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057.