Recognition of multi-modal patterns in full-slide histopathological images using large vision and language models

The growing percentage of cancer patients with the ever-decreasing number of experienced histopathologists creates challenges that can only be solved with the use of new and innovative ML/AI methods. On the other hand, we are witnessing a breakthrough in natural language processing. The creation of publicly available *large language models* (LLM), such as such as BERT, Llama, Alpaca, Generative Pre-trained Transformer (Chat)GPT, etc., have opened up the possibility of relatively easy application in many fields. Even more revolutionary is the use of multi-modal processing, i.e. *both image and text paths*. **We propose to conduct research in order to gain new knowledge and develop new and innovative methods in the field of fusing efficient visual and text encodings for histopathological image classification and diagnosis. It will be able to provide a** *zero-shot* **text response based on scans of histopathological images, as well as the other way around – to present examples of images based on text descriptions. We will explore the possibility of processing medical descriptions in English but also in** *Polish***, which is of great importance for the development of diagnostics in our country. Thanks to this, the proposed system can be used both as diagnostic support for a specialist doctor, as well as a teaching and verification system for young doctors who want to acquire knowledge in this field.** Our project is based on innovative ideas, such as efficient computation of *sparse semantic vision and text encodings*, then their fusion with the *fusing transformer*. Not less important, we'll investigate the possibility of ensuring data privacy using homomorphic encoding, as well as possibility of image content retrieval and improvement of quality of the WSI scans.

The threat of cancer continues to increase worldwide. Malignant tumors are a growing health, social and economic problem around the world. Only in Poland the scale of this problem is determined by the number of new cases, which in 2021 were about 170 thousand and the number of deaths over 100 thousand deaths! Currently over 1,17 million Poles live with cancer. It is estimated that for every 100,000 about 440 people in the Polish population are diagnosed with cancer every year, and over 3,000 have been diagnosed with cancer in the last 10 years. There are many reasons for this condition, including pollution, unhealthy lifestyles, etc. The above facts generate the need to develop research and diagnostic programs as early as possible for cancer diagnosis. The above-mentioned needs can be at least partially fulfilled thanks to modern information technologies, but there is still a need to develop new, effective and safe methods. **The proposed project meets these needs, hence its great scientific and social importance, both in Poland and worldwide.**

In this project, we plan to focus our efforts on prostate and colorectal organs, using data from our previously conducted, as well as other scientific projects. Nevertheless, the previous methods were based only on local patches and allowed diagnosis to a very limited extent. We plan to develop methods of diagnosis based on WSI scans, which primarily use comprehensive information in a form of *vision encodings*. However, even more importantly we propose to join the visual path with the recently developed LLMs. They create wonderful research opportunities, as well as can lead to many revolutionary applications. However, the accuracy and reliability required in histopathology require a special approach, which we propose in this project. General LLMs, such as GPT are not precise enough for the medical diagnosis, including the histopathological one. Therefore the need to develop new and more 'addressed' solutions. To obtain *text encodings*, we will investigate various LLM architectures, suitable for processing and encoding of medical text descriptions, supervised by the specialist pathologist. In particular, we will conduct research and develop appropriate *prompts* information extraction.

The project is planned to be based on the following main tasks:

1. *Research into efficient methods of whole-slide vision encoding.*
2. *Research into fusion of vision models with large language models for visual-language diagnostic system.*
3. *Research into development of novel histopathological data classification methods with data protection.*
4. *Research and development into new methods for histopathological image retrieval with data privacy.*
5. *Research and development into new enhancement methods for the H&E stained whole slide images.*

The project will lead to the development of theories and world-new methods for the classification and diagnosis of histopathological scans and their text description, which will allow for obtaining zero-shot responses and that fulfill data privacy. The project will bring new scientific discoveries and inventions in this field. Due to the great needs in the field of medical diagnostics, it also has great social value.