Assignment 1: Introduction to histopathology image analysis

Project BIA group 1

By: *J.N.A. Bergmans, M. Boone, J.N.A. van Mook, J.F. Peeters*

Exercise 1:

What is the clinical utility of evaluating the presence of metastases in sentinel lymph nodes in breast cancer patients? In other words, how is this information used in the clinical decision making process for breast cancer patients?

Answer:

When one is able to provide prognostic information in the field of oncology, clinical management decisions can be impacted significantly such as treatment and monitoring (Wulczyn et al., 2020)*. Detecting metastases is an important component of surgical managment of breast cancer* (Zahoor et al., 2017)*. More specifically, assessing to which extent the cancer has spread by histopathological analysis of sentinel axillary lymph nodes (SLNs) plays an crucial part in determining the stage of breast cancer* (Ehteshami Bejnordi et al., 2017) *. Lymphatic drainage from all areas within the breast occurs primarily via regional lymph nodes* (Rahman & Mohammed, 2015)*. For many carcinomas, the most common metastatic route is the spreading of the tumor cells via lymphatic drainage of the tumor* (Rahman & Mohammed, 2015)*. Hence the extent to which lymph nodes are involved in breast cancer is considered to be a strong predictor of recurrences and survival* (Rahman & Mohammed, 2015)*. If metastases are detected for example, doctors gain information on the stage of the cancer. Consequently, they might opt for surgery to remove the metastases in order to prevent the cancer from spreading to vital organs. If there is no intervention, the cancer can easily spread further from the lymph nodes to other parts in the body with desastrous consequences for the patient.*

Exercise 2:

How does the introduction of whole-slide imaging change the typical workflow of a pathology lab?

Answer:

The whole-slide imaging technique was the first step towards automation and digitization of analysis tasks previously done by pathologists (Farahani et al., 2015)*. Digitization of metastase analysis allows for less error prone computer assessment rather than humane assessment. It also allows for telepathology, enabling remote interpretation of tissue samples. Over time the technique has had major developments in its usability. Nowadays, WSI is generally approved for primary diagnosis* (Flagship Biosciences, 2020)*. Furthermore, the integration of artificial intelligence will lead to the normalization of computational pathology and inevitably the amelioration of diagnosic accuracy and precision* (Cui & Zhang, 2021)*.*

Exercise 3:

The PatchCamelyon dataset is derived from the CAMELYON16 dataset of whole-slide images. Describe how a neural network classification model trained on small image patches can be applied to larger, whole-slide images with the goal of detecting metastases.

Answer:

From clinical data / images, patient-level labels can be obtained such as developed invasiveness of a lesion or response to a particular treatment (Couture, 2022). *Deep learning algorithms, such as neural networks are able to predict patient labels from histopathology whole slide images images as shown by recent research* (Wulczyn et al., 2020).

NNC models often consist of multiple layers of parameters (or neurons) which are trained by being fed data (Knocklein, 2021). *By feeding a model many instances of normal and metastases images this network 'learns' the ability to recognize metastases. Detection of certain visual characteristics is then initiated by applying the model to large WSI images. However, WSI images can have a size of 100,000 pixels which causes a complete dataset to be too large to fit on a GPU* (Couture, 2022)(Wulczyn et al., 2020).

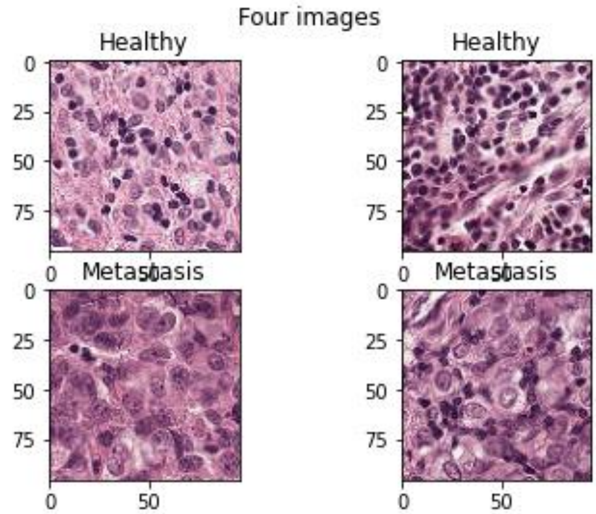
Hence for training a machine learning model, the images are broken into smaller patches. Consequently, feature encodings from multiple patches in a WSI must be aggregated to predict the class. The methods for aggregating these encodings can be used during the training of an end-to-end CNN with a subset of the image patches. (Couture, 2022)

Instead of only aggregating patches, one can also aggregate the patch features. When the aggregation has been completed, fully connected layers can be utilized before the final prediction. Max or mean pooling can be used for feature aggregation to extract a final prediction of the model for the whole image. Moreover, only using the 'most informative' patches or computing a weighted sum can also be used as an aggregation method. (Couture, 2022)

Exercise 4:

Download and unzip the training and validation (4 GB), and testing (1 GB) subsets of the Patch-CAMELYON dataset. Note that the unzipping process might take a while due to the large number of files in the archives.

Write a small Python script that reads and displays a few images from the two classes. Visually describe and compare the appearance of the tissue in the patches with and without metastases.



Answer:

In general, the tissue of the metastases appears to have a higher cell density whereas the healthy tissue seems to contain more open space/ white structures. The cell nuclei in the metastases appear bigger and are surrounded by a smaller cytoplasm in comparison to normal cell nuclei. This makes sense since larger nuclei are often associated with cancerous cells (Jevtić et al., 2014). *Due to the quick cell division of cancerous cells, the cytoplasm is given less time to grow which accounts for the large cell nucleus and relatively small cytoplasm in the cells of the metastases. Besides the size of the nuclei of the cancerous cells, the shape also differs from normal cells. The normal cells have more round, symmetrical smooth nuclei whereas the tumor cells contain nuclei with irregular shapes.*

Exercise 5:

Kaggle username:

biagroup12022

Bibliography

Couture, H. (2022). *From Patches to Slides: How to Train Deep Learning Models on Gigapixel Images With Weak Supervision*. https://towardsdatascience.com/from-patches-to-slides-how-to-train-deep-learning-models-on-gigapixel-images-with-weak-supervision-d2cd2081cfd7

Cui, M., & Zhang, D. Y. (2021). Artificial intelligence and computational pathology. *Laboratory Investigation*, *101*(4), 412–422. https://doi.org/10.1038/s41374-020-00514-0

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J. A. W. M., Consortium, the C., Hermsen, M., Manson, Q. F., Balkenhol, M., Geessink, O., Stathonikos, N., van Dijk, M. C., Bult, P., Beca, F., Beck, A. H., Wang, D., Khosla, A., … Venâncio, R. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, *318*(22), 2199–2210. https://doi.org/10.1001/jama.2017.14585

Farahani, N., Parwani, A. V, & Pantanowitz, L. (2015). Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International.*, *7*, 23–33. https://doi.org/https://doi.org/10.2147/PLMI.S59826

Flagship Biosciences. (2020). *WSI System clearance for primary diagnosis expands clinical application of digital pathology*. https://www.flagshipbio.com/wsi-system-clearance-for-primary-diagnosis-expands-clinical-application-of-digital-pathology/

Jevtić, P., Edens, L. J., Vuković, L. D., & Levy, D. L. (2014). Sizing and shaping the nucleus: mechanisms and significance. *Current Opinion in Cell Biology*, *28*, 16–27. https://doi.org/https://doi.org/10.1016/j.ceb.2014.01.003

Knocklein, O. (2021). *Classification Using Neural Networks - Towards Data Science*. https://towardsdatascience.com/classification-using-neural-networks-b8e98f3a904f

Rahman, M., & Mohammed, S. (2015). Breast cancer metastasis and the lymphatic system. *Oncology Letters*, *10*(3), 1233–1239. https://doi.org/10.3892/ol.2015.3486

Wulczyn, E., Steiner, D. F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., Mermel, C. H., Chen, P.-H. C., Liu, Y., & Stumpe, M. C. (2020). Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLOS ONE*, *15*(6), e0233678. https://doi.org/10.1371/journal.pone.0233678

Zahoor, S., Haji, A., Battoo, A., Qurieshi, M., Mir, W., & Shah, M. (2017). Sentinel Lymph Node Biopsy in Breast Cancer: A Clinical Review and Update. *J Breast Cancer*, *20*(3), 217–227. https://doi.org/10.4048/jbc.2017.20.3.217