**EBIOFY PROBLEM STATEMENT**

This event is an amalgamation of bio-statistics and relevant computational algorithms, namely machine learning, which aims to test budding students on their coding skills, as well as comprehending biological information.



**OVERVIEW:**

Algorithms are to be evaluated for detection and classification of cancer affected regions. This task has high clinical relevance since early detection of anomalies and metastases in affected areas would reduce the workload of pathologists in both detection and diagnosis.

The task in this challenge is to determine a pN-stage for every patient in the test dataset. To compose a pN-stage, the number of positive lymph nodes (i.e. nodes with a metastasis) are counted. There are two categories of lymph node metastasis:

* Macro-metastases: metastases greater than 2.0 mm.
* Micro-metastases: metastases greater than 0.2 mm or more than 200 cells, but smaller than 2.0 mm.

A separate category, called isolated tumour cells (ITC), is strictly not a metastasis, but is rather defined as: single tumour cells or a cluster of tumour cells smaller than 0.2 mm or less than 200 cells. Lymph nodes containing only ITC are therefore not counted as positive lymph nodes.

**NOTE:**

In this challenge, we are using a simplified version of the pN-staging system in breast cancer. The task is to automatically determine per patient which of the following pN-stages applies.

Pathologic lymph node classification (pN-stage):

* pN0: No micro-metastases or macro-metastases or ITCs found.
* pN0(i+): Only ITCs found.
* pN1mi: Micro-metastases found, but no macro-metastases found.
* pN1: Metastases found in 1–3 lymph nodes, of which at least one is a macro-metastasis.
* pN2: Metastases found in 4–9 lymph nodes, of which at least one is a macro-metastasis.

In this challenge we only provide H&E-stained slides, which means that pN0(i-) will not be part of the outcome metrics.

If micro-metastases are found in several lymph nodes, and no macro-metastases are found, that particular case is classified as pN1mi.

For example the patient below, with 5 lymph nodes, should be classified as pN1:

* Node 1: contains no metastases and no ITC.
* Node 2: contains one macrometastasis and one micro-metastasis.
* Node 3: contains two micro-metastases.
* Node 4: contains one micrometastasis and ITC.
* Node 5: contains only ITC.

Possible pN-stages with official definitions: <https://cancerstaging.org/references-tools/quickreferences/Documents/BreastMedium.pdf>

**EVALUATION**

For the evaluation of the results we use five class quadratic weighted [kappa](https://en.wikipedia.org/wiki/Cohen's_kappa) where the classes are the pN-stages. In case of a tie, we rank the submissions by the kappa score calculated on the individual pN stages in reversed order (starting with pN2).

For example if submission A and B reached equal Kappa score, A will be put in front of B if the five class quadratic weighted kappa score of submission A is better than the Kappa score of submission B calculated only on the pN2 stage patients of the ground truth. If it is still a tie we calculate it for pN1 only, and so on.

**DATA SETS FOR PRACTICE:**

<https://drive.google.com/drive/folders/0BzsdkU4jWx9BaXVHSXRJTnpLZU0>

The data in this challenge contains whole-slide images (WSI) of hematoxylin and eosin (H&E) stained lymph node sections.

Depending on the particular data set (see below), ground truth is provided:

* On a lesion-level: with detailed annotations of metastases in WSI.
* On a patient-level: with a pN-stage label per patient.

All ground truth annotations were carefully prepared under supervision of expert pathologists. For the purpose of revising the slides, additional slides stained with cytokeratin immunohistochemistry were used.

**ACCESSING AND ANALYZING THE DATA:**

<https://docs.google.com/document/d/1jTFSTgHQGQ4VXlS-46pAZEWI0EvLMcZavoyKNX7Rhq0/edit>