|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-C-018 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | Lausanne, 23-25 January 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | Institute of Pathology, Charité Universitätsmedizin Berlin, Berlin Institute of Health & Berlin Big Data Center, Berlin Center for Machine Learning, Technische Universität Berlin, Singapore University of Technology and Design, TIL international working group | | |
| **Title:** | | Annotation of histopathology images and benchmarking of AI solutions for the use case "machine learning-based profiling of tumor-infiltrating lymphocytes in breast cancer" | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Frederick Klauschen Institute of Pathology, Charité Universitätsmedizin Berlin Berlin Institute of Health Berlin Big Data Center (BBDC) Berlin Center for Machine Learning (BZML) Germany | | Tel: +49-30-450536-053 E-mail: [frederick.klauschen@charite.de](mailto:frederick.klauschen@charite.de) |
| **Contact:** | | Alexander Arnold  Institute of Pathology, Charité Universitätsmedizin Berlin Berlin Institute of Health Germany | | Tel: +49-30-450536-053 E-mail: [alexander.arnold@charite.de](mailto:alexander.arnold@charite.de) |
| **Contact:** | | Miriam Hägele Technische Universität Berlin Berlin Center for Machine Learning (BZML) Germany | | E-mail: [haegele@tu-berlin.de](mailto:haegele@tu-berlin.de) |
| **Contact:** | | Philipp Seegerer Technische Universität Berlin Germany | | E-mail:  [philipp.seegerer@tu-berlin.de](mailto:philipp.seegerer@tu-berlin.de) |
| **Contact:** | | Alexander Binder Singapore University of Technology and Design (SUTD) Singapore | | E-mail: [alexander\_binder@sutd.edu.sg](mailto:alexander_binder@sutd.edu.sg) |
|  | | Roberto Salgado Chair, International TIL working group Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Université Libre de Bruxelles (U.L.B. Brussels, Belgium | | E-mail: [roberto@salgado.be](mailto:roberto@salgado.be) |

|  |  |
| --- | --- |
| **Abstract:** | This document gives an update on the use case "machine learning-based profiling of tumor-infiltrating lymphocytes in breast cancer" and explains how the histopathology images are annotated and how machine learning models can be benchmarked with the resulting data set of annotated images. The use case was introduced in document [B-014](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-B-014.docx) that contains background information. |

# Introduction

This document gives an update on the use case "machine learning-based profiling of tumor-infiltrating lymphocytes in breast cancer" and explains how the histopathology images are annotated and how machine learning models can be benchmarked with the obtained data set of annotated images. Document [B-014](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-B-014.docx) introduces the uses case and contains background information.

The present document serves as report of work done between meetings (cf. meeting report [B-101](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-B-101.docx)), by summarizing the results of a discussion between pathologists and machine learning experts held on 11 December 2018 at the Institute of Pathology at Charité in Berlin as well as several phone and e-mail discussions as well as weekly group meetings at the Berlin Institute of Health between the medical and ML experts in the subsequent weeks.

Besides, we would like to use this introduction to address two points noted at the Focus Group meeting B at Columbia University (document [B-101](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-B-101.docx)):

1. *"Submitters should be aware that lack of publicly available data will prevent a solution being identified / developed."*

Currently, no high-quality annotated data sets on TILs in breast cancer are publicly available that could be used for evaluation and benchmarking (cf. our document [B-014](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-B-014.docx)) of machine learning approaches. Although some companies, hospitals and research institutes have started to collect digitized histopathology datasets and obtain some level of annotation to train their machine learning models with the goal to develop and explore ways to use machine learning approaches for quantitative diagnostics, these data sets are not made publicly available. Other data public data sets exists, but image and/or annotation quality is limited or annotations by board-certified pathologists do not exist at all. Moreover, even if some of these stakeholders would share well-annotated data, no standardized and "neutral" framework exists to validate ML approaches in an unbiased fashion. With the framework comprising representative and high-quality annotated image data on breast cancer tissue we propose and provide here, the trained models could be benchmarked*.*

1. *"It is desirable to pool subsets of data from different organizations."*

This document explains the annotation procedure and thus enables different organizations to annotate their data in the same and standardized way, which allows for pooling subsets of data from different organizations.

# Annotation of the histopathology images

* Description of the histopathology images that are being annotated.
  + Digitized standard Hematoxylin & Eosin (H&E) stained histological slides
* List/description of the tissue components/classes
  + cancer tissue
    - multiple subtypes
      * focus on NST (no-special-type) and invasive-lobular breast cancer
  + normal tissue
    - normal breast gland and duct epithelium
    - connective tissue (fibers, cells)
    - fatty tissue
    - bone tissue
    - blood and lymphatic vessels
    - nerves
  + immune system
    - lymphocytes
    - granulocytes
    - monocytes/macrophages
    - plasma cells
  + necrotic tissue
  + artifacts
  + background
* Annotations should be flexibly reusable with different patch sizes extractable from annotation coordinates (saved xml-format)
* Annotation procedure (single cell "point" vs. area "region" annotation)
  + positive annotations
    - point annotations (POI): cell nuclei are marked, relevant for heterogeneous tissues (e. g. individual lymphocytes between cancer cells, Fig. 1A)
    - region annotations (ROI): regions containing at least 95% cells of respective class (Fig. 1B/C)
  + negative annotations
    - region annotations (ROI): regions negative of a certain class, i. e. region may contain any cells, but none of the respective class

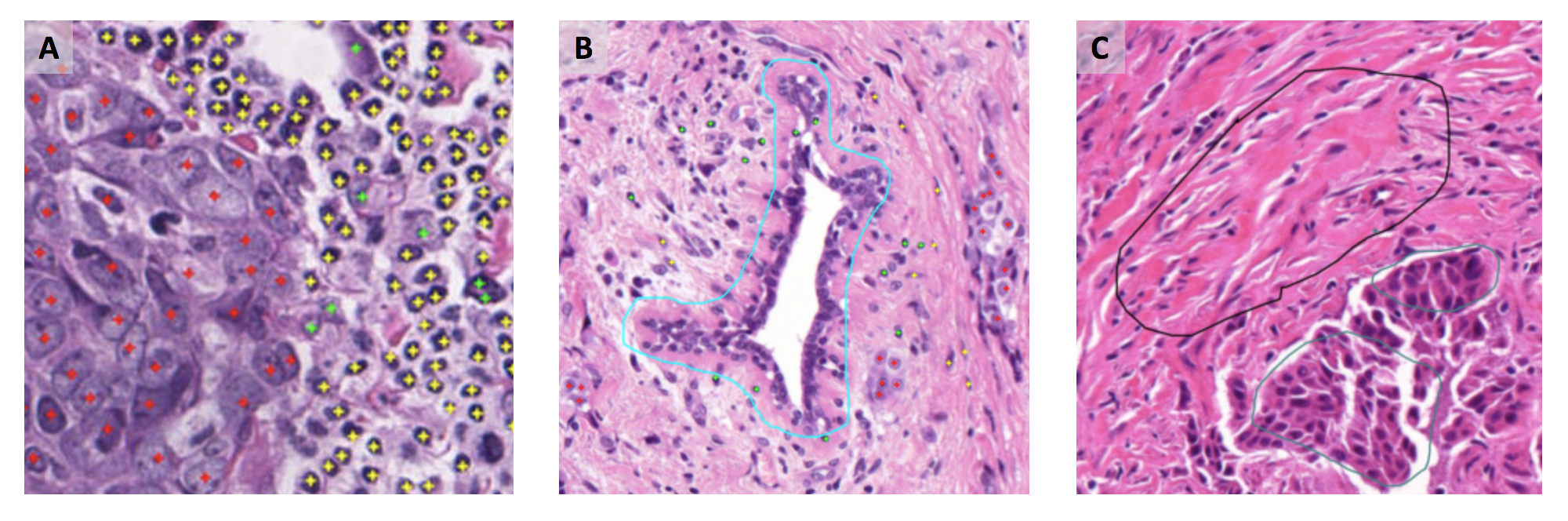


Figure 1. A) Point annotation examples for breast carcinoma cells (red), tumor infiltrating lymphocytes "TiLs" (yellow), fibroblasts (green). B) positive point and region annotations (normal epithelium: cyan). C) negative annotations: negative for cancer (black), negative for lymphocytes (green).

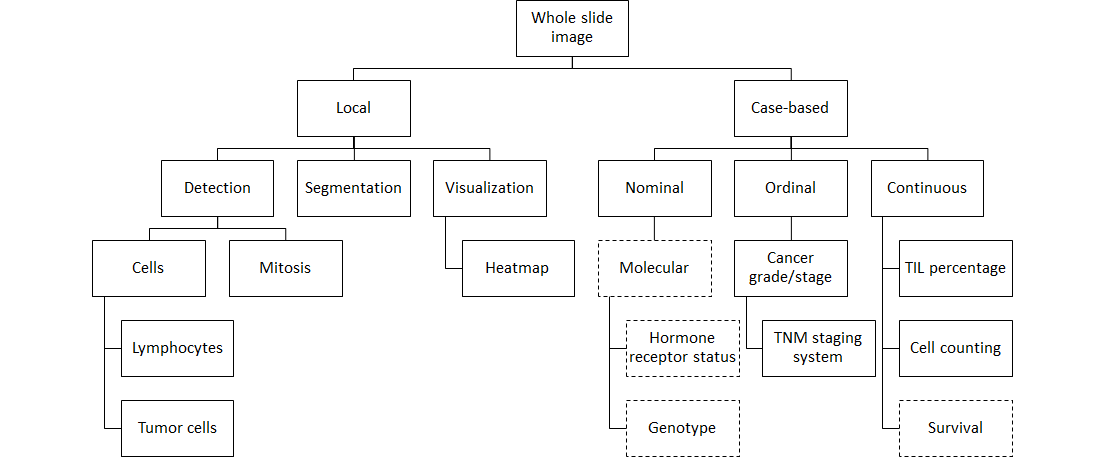
* Description of data and data structures/format:
  + The data will be provided as individual field-of view images (1000x1000 images at 400x resolution in uncompressed tiff-format).
  + Annotations will be available in coordinates of POI and ROI stored in xml-format.
  + 5-10 exemplary images will be made available with test annotations to provide participants with an overview of the data
  + 25-50 densely annotated images will be available in an undisclosed fashion for benchmarking which will be performed on WHO/ITU servers and provide results according to section 3.
  + Large annotated data sets are available at the different participating institutions and the subsets used for the benchmarking will be selected based on consensus by the pathologists to cover a broad and representative spectrum of histomorphological patterns.

# Benchmarking

We expect that several types of machine learning models/algorithms can be benchmarked by our test data set of annotated histopathology whole slide images. With the data, the performance on a range of tasks can be evaluated, as illustrated in Fig. 2. The corresponding benchmarking metrics are listed in Tab. 1.

Local tasks are performed either on image sections, tiles, patches, or pixels, and comprise detection (e.g. of lymphocytes), segmentation and visualization. In case-based tasks, a target variable is inferred from the entire image, and typically belongs to one of the following categories:

1. "Nominal": The target variable is to be assigned/classified to one of a predefined set of classes where the relationship between the classes is irrelevant. In this case, typical performance measures for classification can be used for evaluation.
2. "Ordinal": The target variable still should be classified as in the nominal case but here, the classes are ordered. This should be considered in the evaluation, e.g. a misclassification of a sample belonging to class A into class B should be penalized less than if it was misclassified into class C if the class ordering is (A, B, C).
3. "Continuous": The target is a continuous variable, i.e. the prediction task is regression instead of classification as in the nominal and ordinal case. Therefore, norm-based distance metrics can be readily used for evaluation. Besides, evaluation can be done on pseudo-classes that are defined on intervals of the target variable if small differences are of little importance.



*Figure 2.* Different tasks can be solved with machine learning based on histopathology whole slide images (*top*). Either local "detection" tasks can be performed on image sections/tiles/patches/pixels (*left*) or case-based measures can be computed for the entire image (*right*). (Dashes indicate possible annotations/labels for the future, which are not considered for the present data set yet.)

*Table 1*. Statistical metrics for the performance evaluation on different prediction tasks.

|  |  |  |
| --- | --- | --- |
| **Level** | **Prediction task** | **Metric** |
| Local | Segmentation | Jaccard index, Soerensen-Dice index, ROC/AuC |
| Detection | Accuracy, precision, recall, ROC/AuC |
| Visualization | Pixel flipping |
| Case-based | Nominal target variable | Accuracy, precision, recall, ROC/AuC |
| Ordinal target variable | Accuracy, precision, recall, ROC/AuC,  mean squared/absolute error |
| Continuous target variable | Kaplan-Meier estimator  Comparison to Cox regression  Concordance index  Mean squared/absolute error  k-way accuracy (e.g. long/medium/short survival) |

# References

· Klauschen, F., Müller, K. R., Binder, A., Bockmayr, M., Hägele, M., Seegerer, P., ... & Michiels, S. (2018, July). Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning. In Seminars in cancer biology. Academic Press. <https://doi.org/10.1016/j.semcancer.2018.07.001>

· Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., ... & Treue, D. (2018). Towards computational fluorescence microscopy: Machine learning-based integrated prediction of morphological and molecular tumor profiles. arXiv preprint arXiv:1805.11178. <https://arxiv.org/abs/1805.11178>

· [Standardized determination of tumor-infiltrating lymphocytes in breast cancer : A prognostic marker for histological diagnosis]. Denkert C, Loibl S, Budczies J, Wienert S, Klauschen F Pathologe. 2018 Nov;39(6):520-531. doi: 10.1007/s00292-018-0530-y.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_