**Statement of Work for *Her2* Cancer Imaging**

Version 1.0, 2017-08-22

|  |  |
| --- | --- |
| **Prepared by** | Alex Gordon, Liana Engie, Pia Cortes Zuleta, Victor Moraga |
| **Prepared for** | Mauricio Cerda |
| **Summary of changes** | First draft |

**Background**

Samples were obtained 80 Chilean patients from National Cancer Institute and the gastric cancer biopsy tissues were cut, stained for cell structure and *Her2*, and put onto microscopy slides. Large numbers of slides were scanned by a microscope and loaded onto proprietary software. Individual cells were identified in these images and must be classified into 3+, 2, 1, or 0 based on severity.

**Problem statement**

Goal:

First, the images must be preprocessed to become appropriate inputs into Python. Images are stored in a proprietary file type and are each about 20-30GB 2-dimensional images. To ease calculations the images may also be segmented into individual images.

Next, regions of interest will be identified among the sparse tissue samples. The slides have regions dense with cells (indicated with blue). Cancerous cells are indicated by an outline of brown, indicating where *Her2* is located. The intensity of the color and the percentage of blue/brown cells versus total cells will differentiate the image categories.

Classification of images into multiple groups using machine learning. Different methods will be investigated, including neural networks, to determine which is the most appropriate for categorization of the imaging data into multiple classes. Quantitative data will have to be extracted from the images.

Resources available:

Cancer tissue data and software for viewing

Good internet at all times

High performance cluster

**Deliverables**

A working pipeline in Python to classify images into cancer groups.

**Project timeline**

Wednesday-Friday end-of-day sessions:

Saturday:

Sunday:

Monday:

###

In clinical conditions, gastric samples are treated with specific compounds (HER2), that mark with colors cancer related molecules. A known pipeline [2] is to start from a large 2D image (30 GB per image) to manually look for cells (blue center), surrounded by HER2 marks (brown), and count the number of cancer positive cells (blue center surrounded by brown) vs total cells [2]. As the % of cancer associated cells increases also the cancer “category” (0 to 3).

Milestones:

- Pre-process images selecting non-blank regions

- Computing Regions of Interest (ROIs) associated to cells

- Test supervised approach for classification of ROIs using SVM and CNN with a pre-trained database

- Compare proposed automatic algorithms with clinical available information (known category classification)

Resources:

Literature:

• HER2 testing and clinical decision making in gastroesophageal adenocarcinoma: guideline from the College of American Pathologists. American journal of clinical pathology. 2016; 146 (6), 647-669.

• Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Computerized Medical Imaging and Graphics. 2017; In Press.

Data sets:

• 60 whole slide images from Chilean Patients (from PRECISO clinical study) .

Software:

• Python (numpy, opencv, scikit learn), caffe, NDPI tools

Computing:

• HPC for parallel image processing, GPUs for caffe training.