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

Version 1.0, 2017-08-24

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

This type of test is not a perfect way to detect cancerous tissue, but is much cheaper (and faster, I’m guessing) than other methods. These classifications help practitioners decide which patients require further analysis for diagnosis.

**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 will be cropped into individual smaller tiles.

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.

The main goal is 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**

Current literature research, previous work on this project

Wednesday-Thursday end of day sessions

Research into machine learning algorithms to use

Thursday end-of-day session

Test supervised approach for classification for ROIs using variety of models

Thursday-Saturday

Compare proposed algorithms using available clinical information

Saturday-Sunday

Formalizing of pipeline code

Sunday-Monday

Make presentation

Monday

Give presentation

Tuesday morning

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.