CellMap: Interactive Tissue Analysis

Yidi Ed Liu Sukrut Shishupal

[u1267187@utah.edu](mailto:u1267187@utah.edu) sukrut.shishupal@utah.edu

u1267187 u1401898

Repository link: <https://github.com/sukrut-shishupal/Vis_for_data_science>

1. **Background and Motivation**

The intersection of data visualization and medical imaging presents unique challenges and opportunities that our team is eager to explore. With advancing digital pathology, there is a need for tools that can enhance the interpretability of complex biomedical datasets for clinical and research purposes. We are drawn towards learning the power of technology to solve real-world problems, particularly data visualization to transform how medical data is analyzed and understood.

Histopathological analysis used to rely heavily on manual inspection, which is time-consuming and subject to variability based on multiple observations. We aim to create a visualization tool that assists in the objective analysis of tissue samples and enhance diagnostic accuracy. Our project aims to improve the visualization tools that can lead to better clinical outcomes by enabling precise and faster diagnosis.

Through this project, we leverage our background in computer science and improve our skills in D3.js to contribute to the field of digital pathology by creating a tool that bridges the gap between technical data analysis and practical medical application.

1. **Primary Objectives**

The primary objective of our project is to enhance the analysis of histopathological images through an interactive visualization tool designed to improve diagnostic precision, by taking the model or the prediction of the model, our model should enable users to visualize their model with more control from more dimensions of prospects. Using the tool, pathologists, and researchers can interact with the data, adjusting parameters, zooming, and filtering to examine specific features. It helps in receiving immediate visual feedback, which helps in fostering a learning environment.

We will be using advanced image processing techniques, as it will enhance image quality and feature extraction, revealing patterns and details that are not easily visible in raw images.

1. **Data**

We will be using the Kaggle Breast cancer histopathological image dataset (<https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images/data>). Since the data is on Kaggle, it is publicly available.

1. **Data Processing**

The imaging data will be used to used as the background for comparison, some level of image processing (e.g. image scaling, greyscale adjustment, etc) should be expected. Also, as the imaging data will be used to create our nn model, whose prediction will be the main input to our visualization, how the images are processed will potentially affect our prediction accuracy, thus [] should be expected.

The raw images will undergo a cleaning process, including removing any incomplete filesand standardizing image dimensions and color scales to ensure uniformity across the dataset. To enhance the visual clarity, and quality of the images, color normalization and contrast adjustments will be applied, which will ensure that key features are identifiable. We will then extract key features such as cell size, shape, and clustering patterns using image processing technique, such as segmentation and edge detection. This is help in overlays, to highlight areas of interest in the tissue sample.

1. **Visualization Design**
   1. **Alternative design 1: Prediction over the original imaging**

This will be the most basic design. By taking predictions (with different model parameters), our visualization will appear as a dashboard, where the main window will display the original image as a ‘background’, and the prediction on top of the image. There will also be a control panel area on the dashboard, where the users can select different parameters to display different predictions or select between different images.

* 1. **Alternative design 2: Prediction over the adjusted imaging**

How we process the image will affect our model performance. On the basis of alternative design 1, we will enable another control panel for users to select different [processing], and the predictions as well as the comparative image should change accordingly.

* 1. **Alternative design 3: Statistics**

Just heedlessly browsing through the image list and randomly selecting from it is pointless. To avoid this, we will enable one or more sub-windows in the dashboard to display certain statistics graph(s) (e.g. ROC curve, feature weight, etc) to assist users.

General Ideas:

1) Interactive heatmaps over tissue images.

2) Dynamic filtering of image layers to reveal underlying structures.

Prototype Designs:

1. Basic heatmap overlay.
2. Interactive slider for threshold adjustments.
3. Multi-layer toggling with feature highlighting.

Final Design:

1. Combines the multi-layer toggling from design 3 with the interactive sliders of design 2 for a comprehensive visualization tool.
2. Justification: This design enables detailed analysis while being intuitive for users of varying expertise.
3. **Must-have Features**
4. Accurate heatmap generation.
5. Interactive elements for user-controlled visualization.
6. Responsive design for various devices.
7. **Optional Features**
   1. Self-adjust to different dataset/training target
   2. Api to enableusers with different parameters
8. **Project Schedule**

Week 1-2: Data collection and initial preprocessing.

Week 3-4: Development of basic visualization interfaces.

Week 5-6: Integration of interactive features.

Week 7: User testing and feedback incorporation.

Week 8: Final adjustments and preparation for presentation.