Class Project: Part 2

Developing your program structure and components

As part of your next step towards developing your code, it is important to think about critical steps that your program needs to perform and how to catch possible errors. Evaluation of the project is performed surrounding how successful you achieve the following components for your program. We understand that each program will be different with different data sets and each project will be evaluated as an isolated unit given the constraints of the data set used.

Fill in the following information for your project and upload to your google folder. Inform your point-person when you have completed it and are ready to receive reviews and comments. Keep in mind, your code should have exhaustive inline comments to guide an untrained eye to understand what steps are being performed.

1. **Input Data**: Where is the data coming from: Local/Server? How is the data stored (file format, …). How will you read in the data: user interface / hard coding? In what format should the data be stored within the program: lists, floats, complex, 1/D/2D/3D/nD, etc.. Does the user select which data set/column corresponds to what or is provided within the file as headers. Units?

*These images are my data and some extracted from articles with permissions. They’re going to be stored in either Google Drive or my Documents in my local server. Those images will be stored in folders. As inputs, they’ll be categorized as either Training Data or Testing Data. Those images will usually be in png, tiff, or jpeg. Phase Contrast or fluorescence images are usually in 2D format. The Colab programming will read and extract data from my Google Drive or Documents. Not sure what you mean by “Units?”*

1. **Validate Data**: What types of data are you expecting? Did you confirm no information is provided that is not within the expected range that is physically possible? How do you resolve errors in reading in information: for example, a letter is accidentally present within an otherwise numeric data set. Are units appropriate for anticipated data set?

*The initial data I just images. The output will be 1) images with chosen, plotted equations for analysis and 2) texts containing the analysis of those images with chosen, plotted equations. I don’t understand what you mean by “*Did you confirm no information is provided that is not within the expected range that is physically possible?” *There will be two kinds of errors: reading the data and the methods of analyzing the data. For reading the data, most of them will be the usual coding error you see in Colab such as “Invalid syntax.” The analysis errors will be covered in Visualize Results and Export Results because they can contain valuable information, e.g. statistical outliers. I still don’t understand what you mean by “Units” in “*Are units appropriate for anticipated data set?”

1. **Manipulate/Analyze Data**: What operations need to be performed to analyze the data set and obtain the desired results. Are reference values needed which also need to be read into the script as a separate file? Is the analysis properly translated into an algorithm? Does the algorithm fail gracefully if an error is encountered (unphysical output)?

*The goal is to make my computer analyze images. Human scientists analyze images by guesses and approximations most of the time. While a computer uses approximations, it’s based on mathematical equations and fit approximations rather than just “looking at them.” Their greyscale mathematical analysis might be better than how human eye views color lines. The general algorithm would be just reading the cell’s membrane outline based on greyscale or colorscale borders between the cell and the background, and then choose an equation that matches this borderline pathway to determine cell shape (each side individually or altogether). This equation will then compare to the borderline to see if they best fit and how much by correlation (R2 for example). From that, the program will export out a text that’ll outline 1) which equation used for each image, 2) how well it correlates to the cell membrane’s greyscale or colorscale borderline and 3) (assuming high correlation) some analysis on cell viability, adhession, and which process (proliferation or differentiation); otherwise, just “Error, no equation would best fit here.” That error would be either 1) need to search other articles with better fit equations or 2) a scientific barrier to point out. Either way, that error is more on intellectual debate and research progress rather than coding issue. No need for reference values here so far.*

1. **Statistical Analysis**: What statistical analysis can be performed on the data set to evaluate errors, correlations, etc.? Do you need to fit a function to a data set?

*Just fitting mathematical equations onto cell membrane’s pixel curve to see 1) if there’s good correlation between those curves and the equation used and 2) to determine viability, adhesion level, and the function (proliferation or differentiation) at that moment.*

1. **Visualize Results**: How do you best represent your results? What type of visualization are appropriate and effective? Show the raw data to the user along with your analyzed data. If you took a class on data visualization, make sure to adhere to best practices for colors, plotting style, labeling, etc.

*The original images will still be stored in Training Data and Testing Data. The new image, containing chosen, plotted equation, will be in the new folder: Results. In addition, a text analysis for each of those images will be saved in that folder (Google Drive or Documents in Local Drive).*

1. **Export Results**: How will you export the data? What should be exported? What file formats are appropriate for the user to perform additional analysis or incorporate the data in an effective way in a presentation or paper? How do you ensure the user knows how to link your calculated errors with the original data set? Where will you export the results? Make sure to export both numeric/text information **and** your visualizations (images, plots, etc.).

*Those new images and texted analysis will be in Results folder on Google Drive or Documents (Local Drive). The images will all be in the same format. The texted analysis can be in Word Doc. or just Text. The images in Result folder should have plotted equations used for reading cell membrane’s outline and a text analysis of that. That image and its text analysis should have the same name to avoid confusion.*