Project Proposal

Project 13, **Counting stomata cell structures in plant leaves**

Team Members:

Sihan Liu
Wen Yan
Xingyu Pan
Ruilun Wang
Zhengyu Wu

Motivation (Why?):

Leaves are the photosynthetic organs of plants and have a major influence on plant growth. Meanwhile, gas exchange and transpiration are regulated by cells within the leaves surface, stomata.

The stomata are the key factor to control the growth of plants, it can open on the surface of leaves, to allow Carbon Dioxide into the leaf for photosynthesis, or can close to prevent water loss when conditions are hot and dry. Since its special function, the stomata characteristics can be used to determine a plant's growing speed and its needs for the surrounding environment. Therefore, it is important to have a tool that can help researchers to have enough detailed information for their research on the relationship between plant growth and their stomata traits.

Project Goal and Research Questions (What?):

In order to study how leaf stomata properties are genetically controlled in poplar trees, and in turn, how these traits influence how the trees grow, we aim to design a program to help researchers analyze the stomata by scanning genomically characterized images. Through this program, we will analyze stomatal traits, including stomatal density, stomatal size, and pattern of stomata. By collecting relevant data on stomata, we will research how various stomatal traits affect plant growth. Find the size, density, and or pattern of stomata that correlate with the growing speed and height of the tree. Also, we want to analyze the commonality and difference between different stomatal traits and explore correlations between these traits.

With the advance of this project, we believe that we will put forward more reasonable and constructive research questions.

Data: Description, Sources, etc. (How?):

The data may have the following format:

Feature 1:
A microscopic image of the leaf cells with stomata

Feature(s) 2 ~ j:
Other properties not included in the image

Dependent Variables:
Stomata number (density), size, and pattern

The dependent variables about stomata density and size may be continuous, while its pattern may be defined as qualitative.

Sources:
Datasets provided by Dr. Groover

Methodology (How?):

- **Image processing and data gathering**: from the leaf stomata images, we can observe many characteristics of stomata, including size, density, and pattern. By using image processing tools in python, we are able to improve the quality of the original images and have more flexibility while using them. In addition, the data set may contain other properties other than images. Hence, it is necessary to apply the data cleaning method to format the data with the correct type before we can make further progress on comparing and analyzing.
- **Exploratory data analysis:** we would like to explore the data sets before we start on our models. This is useful for generating some interesting hypotheses or research questions that cannot be overlooked.
- **Usage of Stomata Counter**: to improve the efficiency and accuracy of stomata identification, we can upload the leaf epidermal micrographs to an online AI tool named “Stomata Counter” and get the result dataset containing stomata density, size, and other information in a short time.
- **Modeling (supervised & unsupervised learning)**: given any input features, our goal is to predict the dependent variables using regression and classification models. Following the tradition, we divide the data into a training set and a testing set, one for training and the other for testing our models. When the models reach some performance level, we would then use unsupervised learning techniques to cluster the data (including the dependent variables as inputs) and discover some association rules among the inputs.

Work Tasks:

- Data Gathering
    - Collect dataset from the source
    - Data cleansing (programming is included here)
    - Package preprocessed data
- Programming
    - EDA
    - Build models
    - Train and test their performance
    - Plot graphs (Mainly about model statistics)
- Analysis & Research
    - Analyze density related to other variables
    - Analyze patterns related to other variables
    - Clustering (also include programming here)
    - Plot graphs (Mainly about data visualization)
- Writing
    - Progress report
        - Weekly reports
    - Final report
        - Scientific paper

Distribution to Group Members:

- Data Gathering
  Xingyu Pan
  Zhengyu Wu
  Wen Yan

- Programming
  Ruilun Wang
  Xingyu Pan
  Sihan Liu
  Zhengyu Wu
  Wen Yan

- Analysis & Research
  Ruilun Wang

Wen Yan
Sihan Liu

- Writing
  Ruilun Wang
  Zhengyu Wu
  Sihan Liu

Timeline:

* All dates mentioned in this part are tentative due dates.

- Milestone 1:
  - Project Proposal
    - First Draft: Oct 23
    - Final Draft: Oct 27
- Milestone 2:
  - Data Gathering
    - Collect dataset from source: Oct 30
    - Exploratory Data Analysis: Nov 3
    - Package preprocessed data: Nov 3
- Milestone 3:
  - Models: Nov 22
    - Building and training the models
    - Testing models
- Milestone 4:
  - Model improvement:
    - Improving Model Performance: With the project progress
  - Analysis & Research: Dec 6
- Milestone 5:
  - Writing:
    - Weekly Report: Due weekly
    - Final Report: Dec 15

Final Due Date: Dec 18.

Mentor's comment:

“I would suggest that you limit the scope of your proposal. First, I would choose either leaf shape, or else stomata patterning but not both. Second, I would suggest that you not propose to do a genetic/genomic analysis, as this is going to be quite complicated and I am not sure you would be able to do all this in the time you have allotted.

I can send you information about either stomata or leaf shape analyses. Again up to you but I would not suggest doing both.”

--- by Dr.Groover’s email.