Classifying Physiological States of Hair-Cap Moss (*Polytrichum*) Based on Digital Images Using Machine Learning

YUAN GAO CHRIS FERNANDES STEVE RICE (ADVISORS)
COMPUTER SCIENCE DEPARTMENT, UNION COLLEGE

Introduction, Motivation, and Background

Mosses have an important influence in the environment, including productivity, thermal environment of the soil, nutrient cycling in the ecosystem and even attach influence on other plants in the environment [1]. Hair-cap moss is desiccation tolerant. It can withstand drought-like conditions and return to normal growth state after rehydration. The physiological states of mosses changes during this dehydrating and rehydrating process. The physiological features and growth rates of mosses differs in different states. Being able to monitor the physiological states using near-surface remote sensing of hair-cap mosses will provide valuable information for predicting the growth of mosses and assessing the vegetation condition in boreal forests. The initiative is to classify the physiological states of the mosses based on digital images of moss canopies since such digital images are easy to acquire.

Question

Can we find an algorithm to identify hydrated areas from desiccated areas using only digital images of the moss canopy?

Methods

- Collect digital images of hair-cap mosses in the field: 196 in hydrated state, 200 in desiccated state and 50 in mixed state
- Images taken 25 cm above the ground surface and covered a 24×18 area
- Use OpenCV and ImageJ program to process images and extract image information attributes based on color and the structure of shoots
- Build dataset with attributes extracted from images of moss canopy surface
- Use Weka Machine Learning Library to identify ideal machine learning algorithms

Results

- Information extracted from images (ranked by contribution to the model):

1. Percentage of Green Area Coverage
2. Circularity
3. Average Size of Particles
4. Number of Particles
5. Average Aspect Ratio of Particles

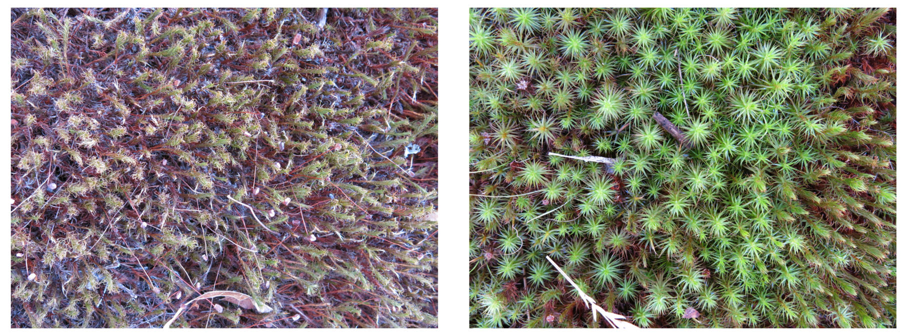

FIGURE 1: Sample images of hair-cap mosses (left: hydrated right: desiccated), outside Cooperstown, NY.

- kNN classification algorithm has the highest classification accuracy and performs statistically significantly better than other algorithms.

Algorithms	Percent Correct Classification
ZeroR	33.56
OneR	73.17
JRip	92.10
J48	92.45
kNN	95.25
Naive Bayes	88.54

TABLE 1: Classification accuracy of selected algorithms with 10 folds cross-validation.

Evaluation

- Manually classify the physiological states through a MATLAB script

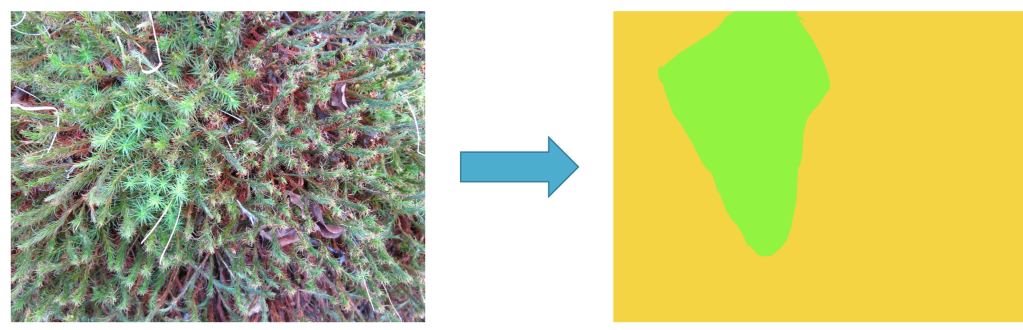

FIGURE 2: Sample input (left) and output (right) image of the MATLAB program.

Evaluation

- Map kNN model prediction decisions onto the original image

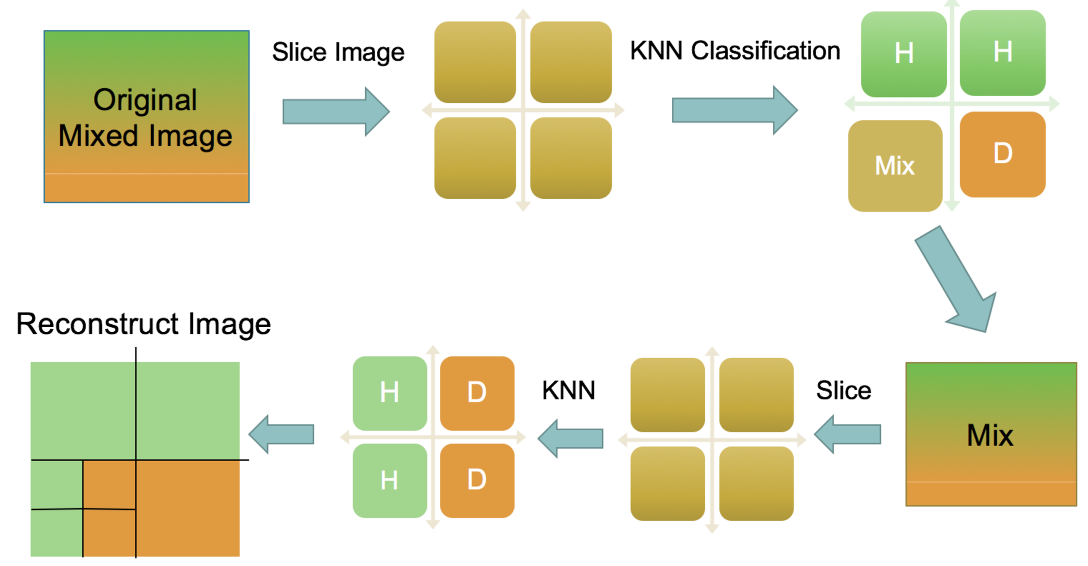

FIGURE 3: Visualization of the pipeline process

- Compare kNN classification with manual classified images

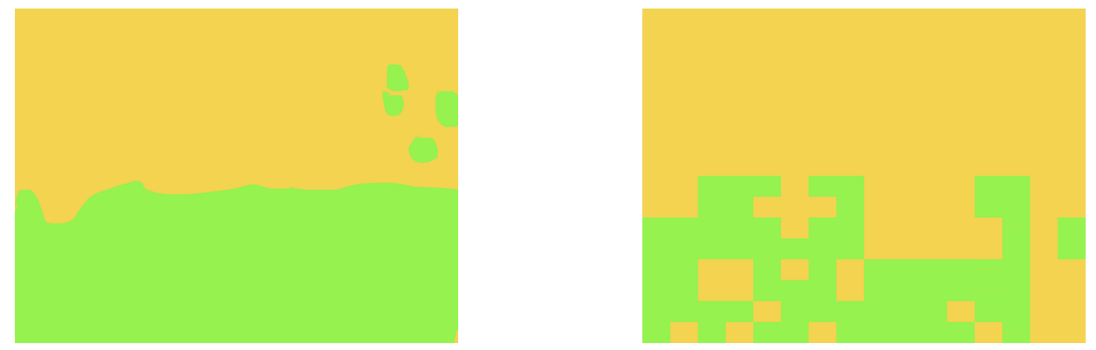

FIGURE 4: Sample pipeline output. Left: manual classification Right: assembled image based on model prediction.

- Overall accuracy of images generated by the pipeline

Average area classified correctly: 66.38%
Percent hydrated area classified correctly: 34.75%
Percent desiccated area classified correctly: 79.52%

Acknowledgments

Special thanks to Prof. Steve Rice, Prof. Nick Webb and Christina Dykas ('18).

References

[1] OECHEL, W. C., AND VAN CLEVE, K.
The Role of Bryophytes in Nutrient Cycling in the Taiga.
Springer New York, New York, NY, 1986, pp. 121–137.