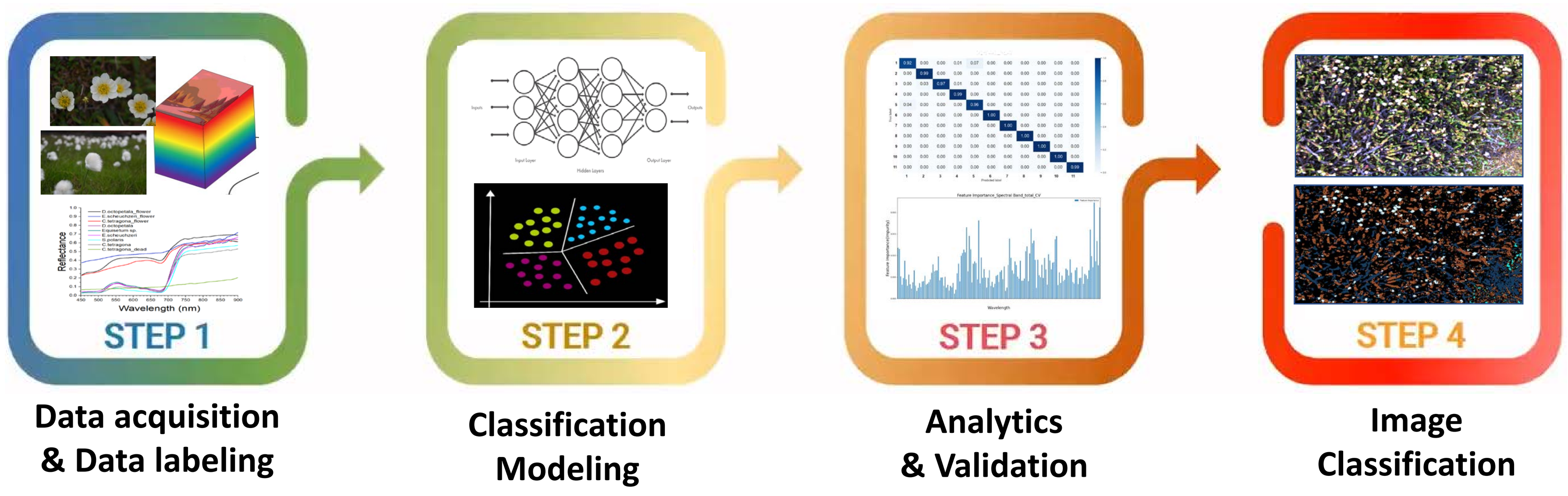
Machine Learning-based Classification for mapping Arctic plant species using Terrestrial Hyperspectral Imagery

Junyoung Yang[p1,2], Yoo Kyung Lee[1,2] , Junhwa Chi[c3]
[1]Korea Polar Research Institute, Incheon 21990, Republic of Korea, [2]University of Science and Technology, Incheon 21990, Republic of Korea
[3]Major of Big Data Convergence, Division of Data and Information Sciences, College of Information Technology and Convergence, Pukyong National University, Busan 48513, Republic of Korea
***Corresponding author: jhchi@kopri.re.kr**

Summary

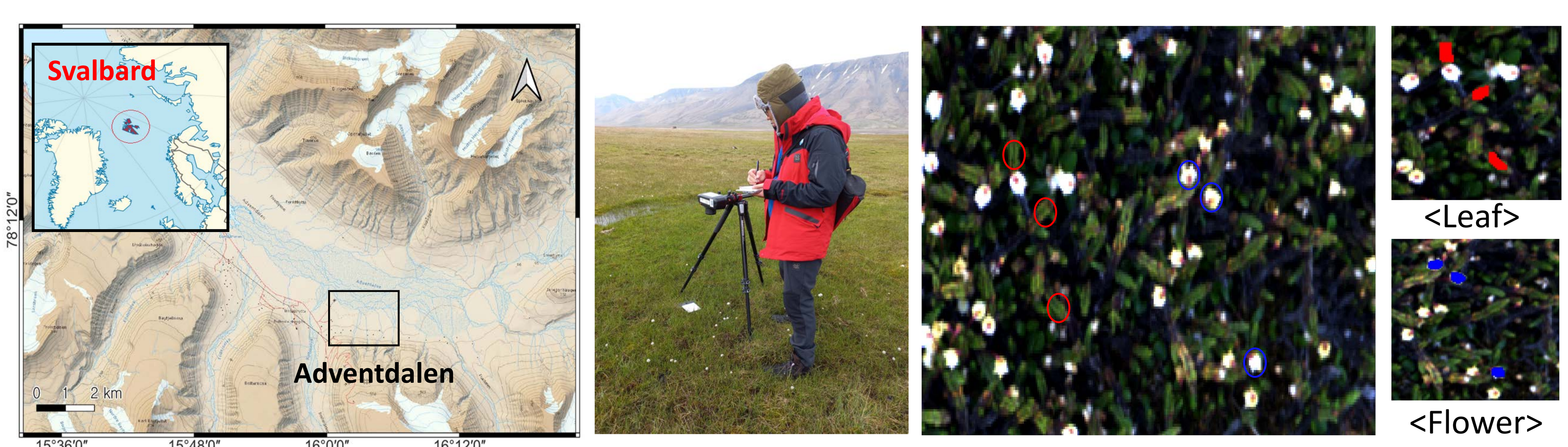


Introduction

- Hyperspectral remote sensing is a valuable tool for monitoring the rapid changes in the distribution and composition Arctic vegetation in response to climate change.
- However, the potential of hyperspectral imagery for mapping Arctic plant species remains poorly established due to the difficulty in data acquisition from Arctic region.
- To address these limitations, this study included two research purposes, (1) collecting hyperspectral information for Arctic plant species, (2) investigating the mapping potential of the hyperspectral imagery through classification experiments.

Methods and Materials

1) Data acquisition

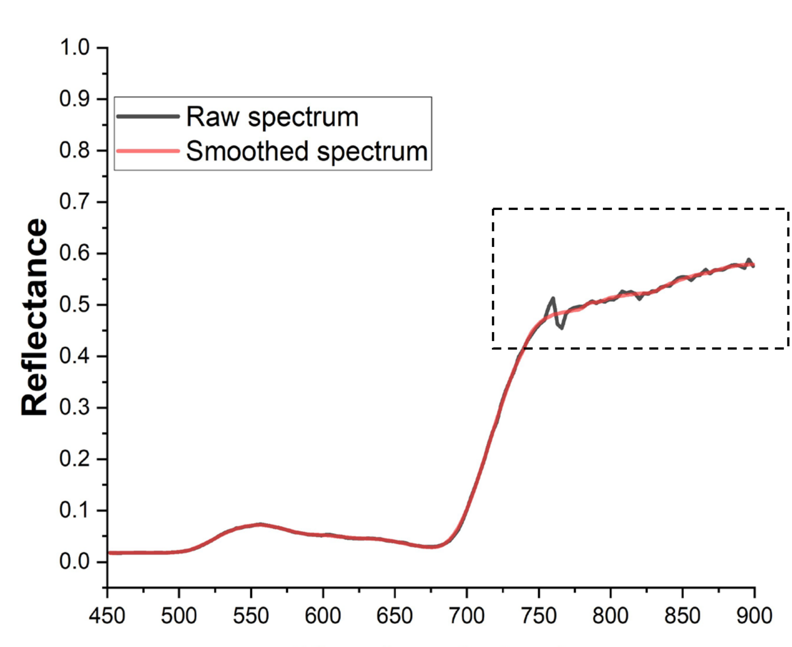


- Hyperspectral data were collected using a terrestrial hyperspectral camera (Specim IQ) in Adventdalen, Svalbard (78˚09′55″ N and 16˚00′00″ E).
- Hyperspectral pixel data associated with Arctic plant species were manually extracted from the hyperspectral images and were labeled using visual inspection.

2) Data preprocessing

2-1) *Spectral smoothing*

- To address noise issue, the Savitzky-Golay filtering was applied to the raw spectrum to obtain the smoothed reflectance spectrum.



2-2) *Spectral derivative*

- We investigated whether the classification performance could be improved by applying the spectral derivative technique.

2-3) *Spectral similarity*

- Spectral correlation coefficient (SCC) was used to evaluate the spectral separability of Arctic plant species.

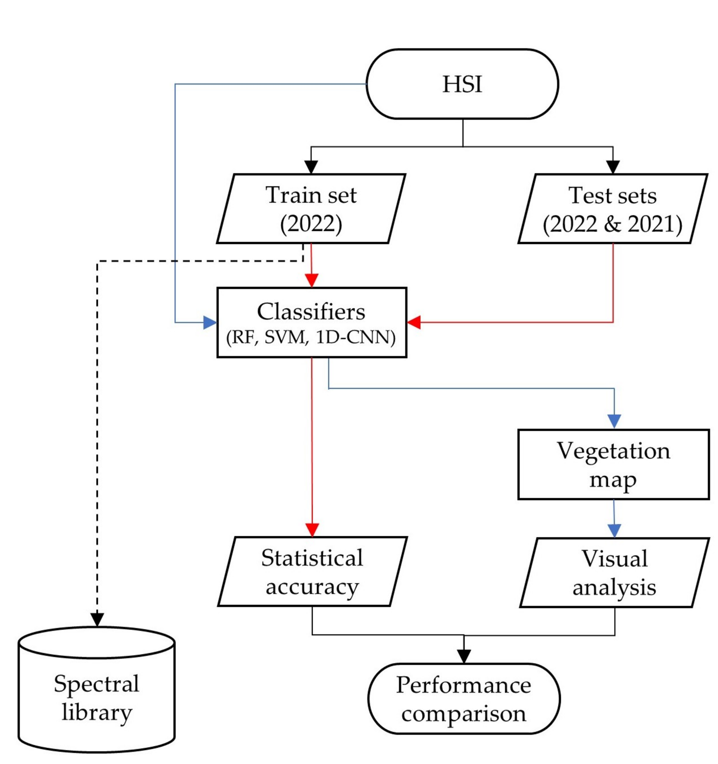
➢ The **first derivative** and **SCC** were calculated using the follow equations, respectively. ($dS/d\lambda$ = First derivative spectrum, S = reflectance, λ_i = Wavelength of i-th band) (n = the number of spectral bands, Two spectra, $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$)

$$\frac{dS}{d\lambda} = \frac{S(\lambda_{i+1}) - S(\lambda_i)}{\lambda_{i+1} - \lambda_i}$$

$$SCC = \frac{n\sum_{i=1}^{n} a_i b_i - \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i}{\sqrt{\left[n\sum_{i=1}^{n} a_i^2 - \left(\sum_{i=1}^{n} a_i\right)^2\right]\left[n\sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} b_i\right)^2\right]}}$$

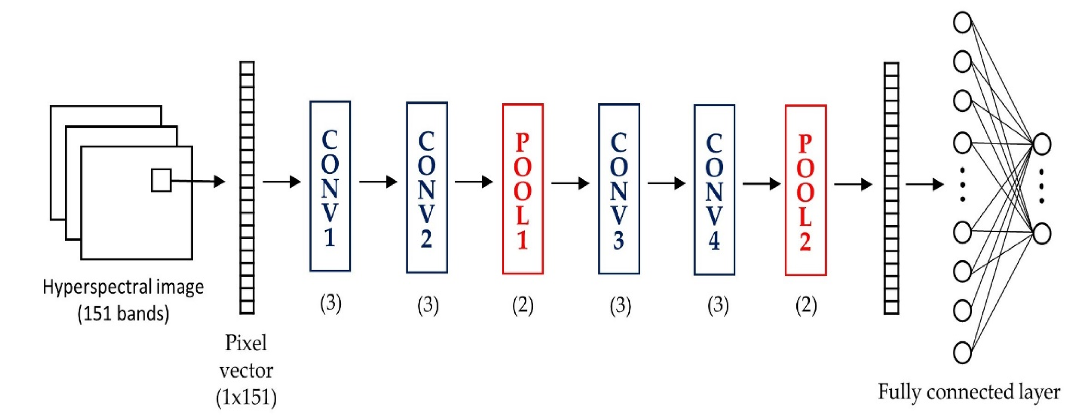
3) Research flow

- Labeled datasets were divided into three groups, training set (2022) and test sets (2022 and 2021).
- Classifiers were trained using training set (2022) with hyperparameter tuning.
- Classification performance of each classifier was evaluated using two test sets (2022 and 2021).

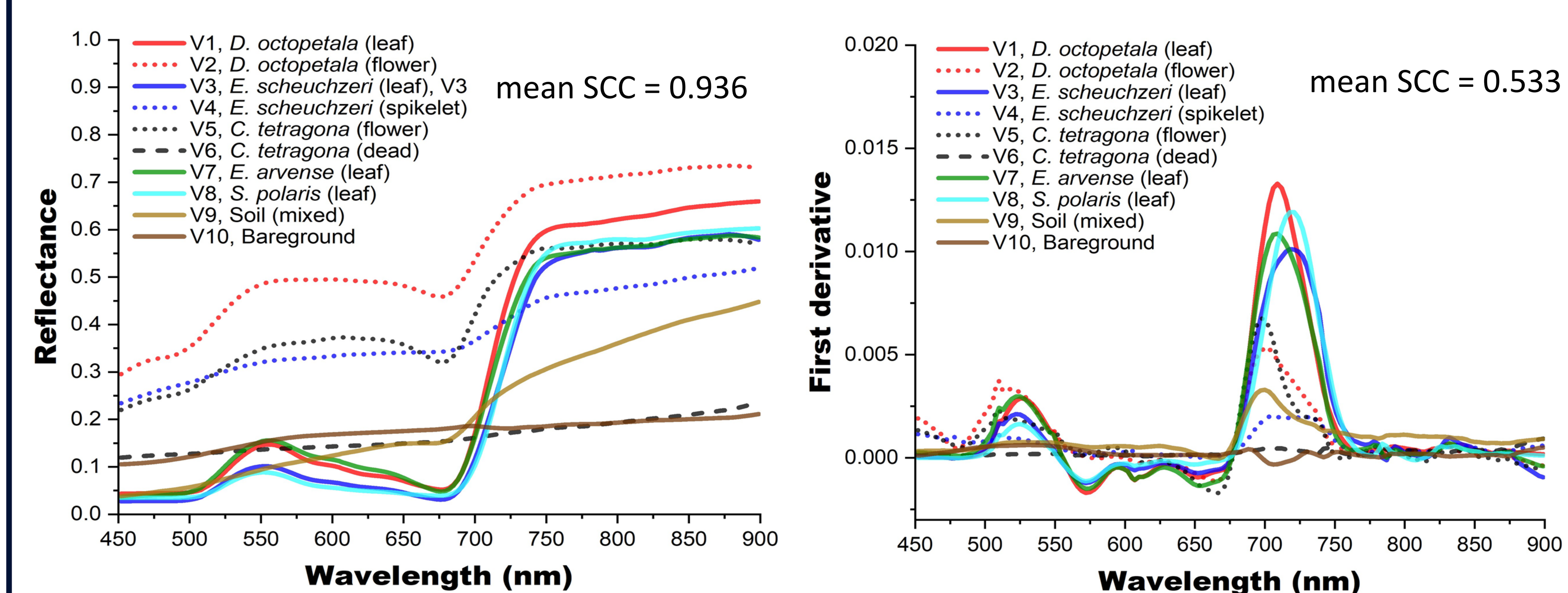


Machine Learning-based Classifiers

- The performances of three representative classifiers, Random forest (RF), Support vector machine (SVM), One-dimensional convolutional neural network (1D-CNN), were compared to determine the optimal classifier.
- In the architecture of 1D-CNN, the rectified linear unit (ReLU) activation function was used for CNN layers, Softmax activation was applied to the output layer.



Spectral Library



- Spectral patterns of the classes were visually distinguishable in the first derivative.
- Applying the first derivative resulted in a decrease in the mean SCC, indicating an increase in spectral similarity, compared to the reflectance spectral library.

Comparison of Classification Performances

1) Statistical Accuracy of Classification performance

<Reflectance>

Dataset	Classifier	Accuracy
2022	RF	91.50%
	SVM	96.41%
	1D-CNN	97.24%
2021	RF	74.28%
	SVM	89.04%
	1D-CNN	92.80%

<First derivative>

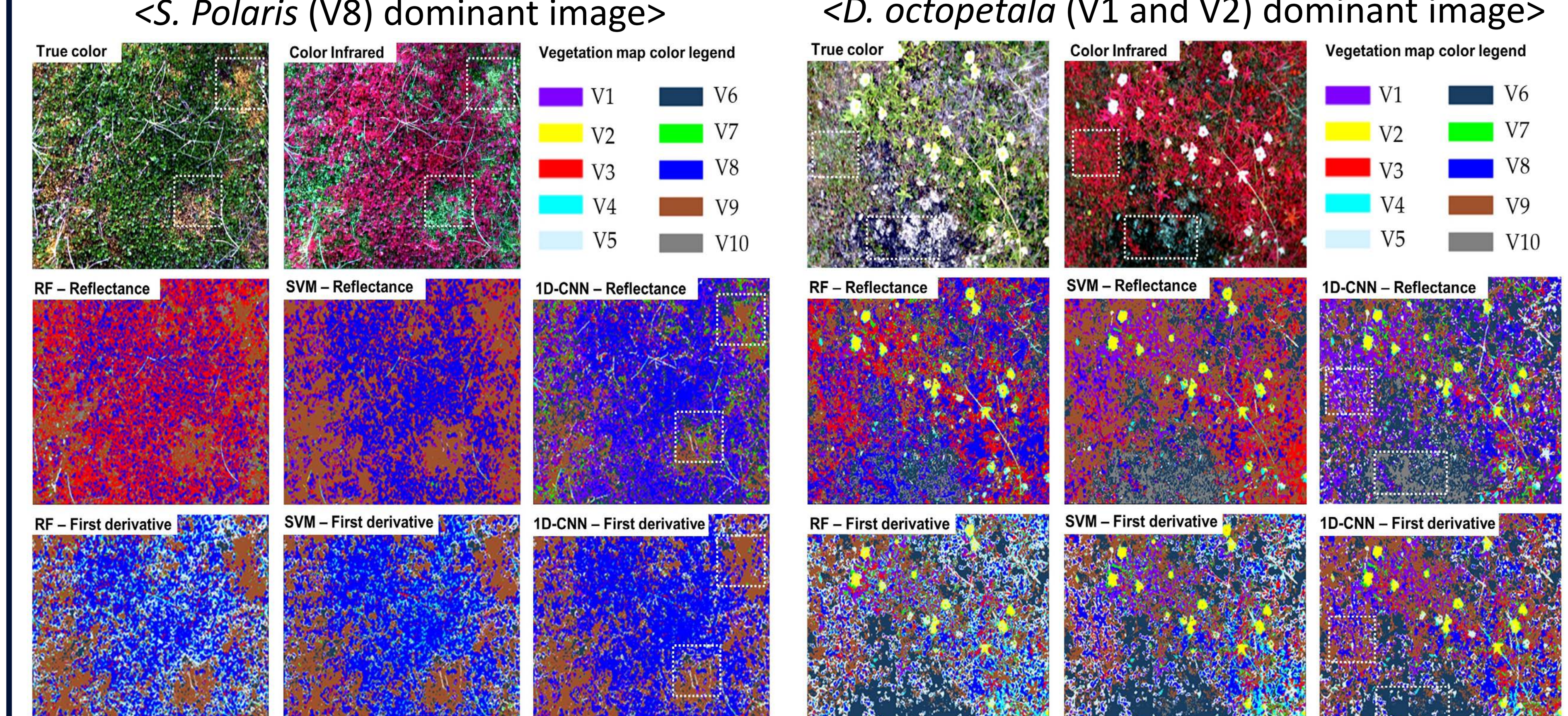
Dataset	Classifier	Accuracy
2022	RF	96.60%
	SVM	98.52%
	1D-CNN	98.88%
2021	RF	90.64%
	SVM	87.54%
	1D-CNN	93.52%

➢ The 1D-CNN classifier combined with the first derivative technique showed most stable performance.

➢ It achieved the highest statistical accuracies for two different test sets (highlighted in yellow).

2) Visual Inspection of Classification Results

<S. Polaris (V8) *dominant image>* *<D. octopetala* (V1 and V2) *dominant image>*



➢ The 1D-CNN with the first derivative produced most accurate vegetation map, showing better visual agreement with true color and color infrared images (white dotted rectangles) (V1 – V10 classes were delineated in the spectral library.)

Conclusion & Future Study

- The hyperspectral data collected in this study, including the spectral library, provide a valuable reference for understanding the spectral characteristics of Arctic plant species.
- 1D-CNN with the first derivative is a promising approach to enhance spectral separability of Arctic plant species, contributing to the monitoring Arctic terrestrial ecosystem.
- We plan to expand our study to UAV hyperspectral analysis for large-scale mapping in the Arctic, and to conduct remote monitoring studies of plant species.

*Funding: Korea Polar Research Institute(KOPRI) granted by the Ministry of Ocean and Fisheries (KOPRI PE22450, PE22650).