

Exercise 9: Supervised Classification

50 Points scaled to 20 Points

Introduction

In this exercise, you will undertake a supervised classification of a Sentinel-2 Multispectral Instrument (MSI) multispectral image collected on 7/29/2020 over the city of Vancouver in British Columbia, Canada using ArcGIS Pro and the Image Analyst Extension. You have also been provided with randomized validation points to create error matrices and assess and compare models. In this exercise, you will specifically compare the support vector machine and random forest machine learning algorithms.

Objectives

- *Interpret multispectral imagery to create training data*
- *Use training data, input imagery, and machine learning algorithms to create land cover classification models*
- *Use trained models to predict images and create land cover predictions*
- *Assess and compare classification performance using classification results and randomized training data*

Prerequisite Materials

- ❖ Modules: Classification and Accuracy Assessment
- ❖ Videos
 - Lab 9 Intro: <https://youtu.be/tytnxVwpxgc>
 - Supervised Classification: <https://youtu.be/SZbLwoyXTSw>
 - Confusion Matrix: <https://youtu.be/6XHyYh45PxA>

Data

- ❖ **s2_2020_7_29_vancouver.tif**: Sentinel-2 image of the city of Vancouver and surrounding area in British Columbia, Canada. The bands have been stacked into a single, multiband image. See the table below for additional information. Note that the 60 m bands were not included in the stack (Coastal Aerosol, Water Vapor, and Cirrus Cloud). All 20 m bands were resampled to 10 meters.
- ❖ **validation_points**: point samples for validating final classification products. The “GrndTruth” field represents the correct classification. When you add classification results, you will need to name the new field “Classified”.

Note: The Alexandria map was provided to show some example training data and classification output. The image represents a Sentinel-2 MSI scene from 8/4/2020. Four classes were differentiated: Developed, Sand/Rock, Vegetation, and Water. Two results were generated, one using random forest and the other using support vector machines. The results here are similar to those you will obtain for the Vancouver image.

Note: The Topo map provides example results for a classification of mine features from historic topographic maps. Using the sample points and results, a confusion matrix was generated (C_0 = Not Mining and C_255 = Mining). This provides an example confusion matrix, similar to the ones you will obtain for the Vancouver map. The classification was generated using deep learning, so is pretty accurate.

Band	Included	Band Number in Stack	Native Spatial Resolution
Band 1 (Coastal Aerosol)	No	NA	60 m
Band 2 (Blue)	Yes	1	10 m
Band 3 (Green)	Yes	2	10 m
Band 4 (Red)	Yes	3	10 m
Band 5 (Red Edge 1)	Yes	4	20 m
Band 6 (Red Edge 2)	Yes	5	20 m
Band 7 (Red Edge 3)	Yes	6	20 m
Band 8 (NIR)	Yes	7	10 m
Band 8A (Narrow NIR)	Yes	8	10 m
Band 9 (Water Vapor)	No	NA	60 m
Band 10 (Cirrus Cloud)	No	NA	60 m
Band 11 (SWIR 1)	Yes	9	20 m
Band 12 (SWIR 2)	Yes	10	20 m

Background Questions

Question 1. Explain the difference between supervised and unsupervised classification. (4 Points)

Question 2. Explain the difference between spectral and informational classes. (4 Points)

Question 3. What are two benefits of using machine learning algorithms (e.g., random forests, support vector machines, k -nearest neighbors, etc.) as opposed to traditional methods (e.g., minimum-distance-to-means, parallelepiped, and maximum likelihood) for supervised classification tasks? (4 Points)

Question 4. Why is it important that validation data are randomized? (4 Points)

Required Steps

To complete this exercise, you will need to do the following:

- ❖ Open the project file.
- ❖ You may want to experiment with changing the band combinations. See the table above for the layer designations. The image is currently displayed as Red = SWIR 2, Green = NIR, and Blue = Green.



Figure 1. Sentinel-2 image used in lab exercise.

- ❖ Create training samples using the Training Sample Manager. See the explanations of the classes provided at the end of this document. You should collect a minimum of 20 samples for each class. Make sure to capture the spectral variability of each class.
- ❖ Use the image, training samples, and Train Support Vector Machine Classifier/ Train Random Trees Classifier tools to train the support vector machines and

random trees (a.k.a. random forests) classifiers and obtain classification maps. You can use the default parameter arguments.

- ❖ Use the resulting models and the Classify Raster Tool to classify the image and obtain two separate classifications.
- ❖ Use the two classification results and the provided validation data to assess and compare the models. You can extract the class codes using the Extract Multi Values to Points Tool and create a confusion matrix using the Compute Confusion Matrix Tool. See the linked videos for a demonstration of this process.

Question 5. Provide both of your resulting error matrices. (4 Points)

Question 6. Which classifier yielded the highest overall classification performance based on overall accuracy? (4 Points)

Question 7. Which classifier yielded the highest overall classification performance based on the Kappa statistic? (4 Points)

Question 8. Based on the user's and producer's accuracies, which classes were most difficult to distinguish? (4 Points)

Question 9. Based on the error matrices, what classes were most confused? (4 Points)

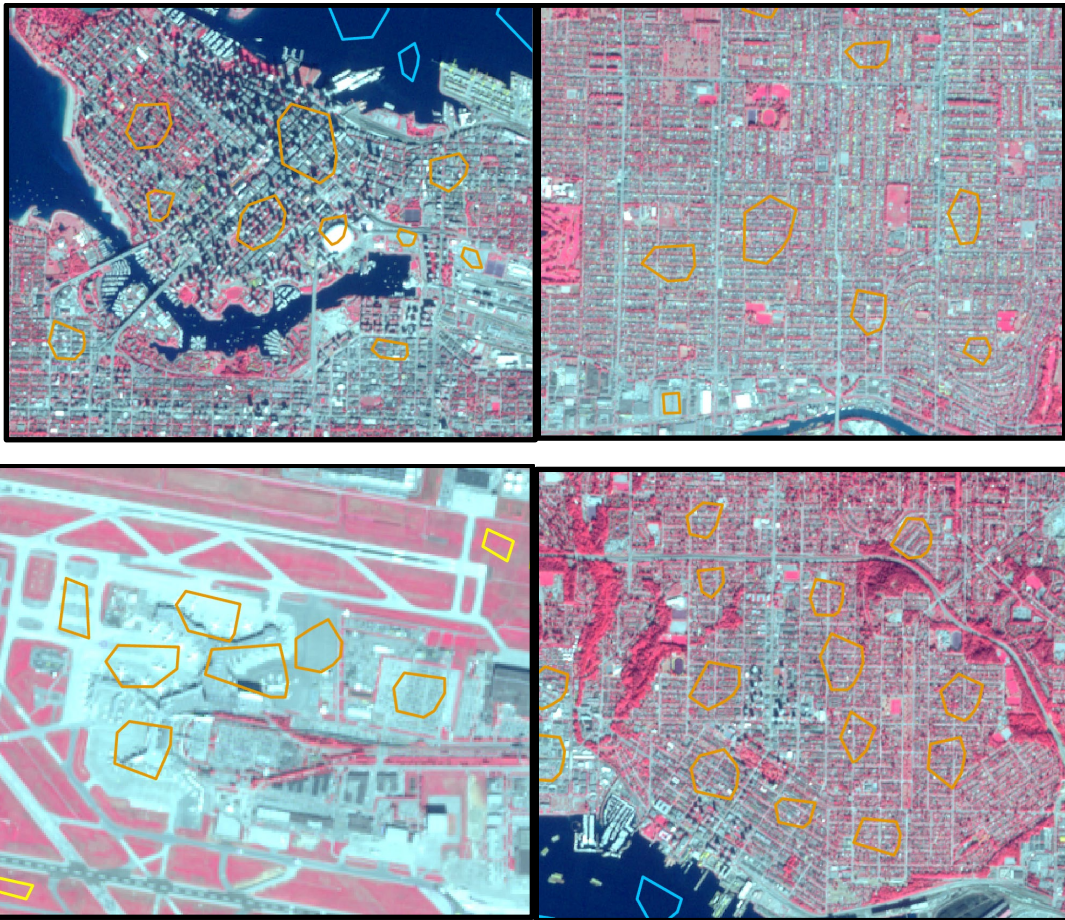
Question 10. List and explain two methods that could be explored to potentially improve the classification results. (4 Points)

Deliverable 1. Create a map layout that includes 4 separate map frames that show (1) the input imagery, (2) your training samples, (3) your support vector machines classification result, and (4) your random forest classification result. Make sure to include a legend explaining the land cover maps. The layout will be judged based on neatness, presentation, and use of space. (10 Points)

Cover Classes

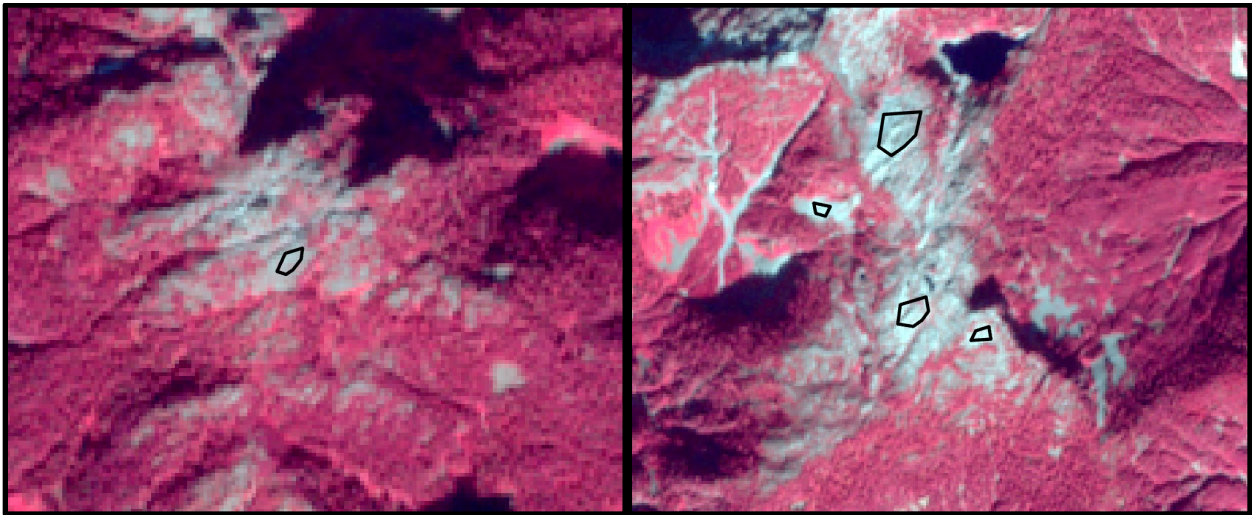
Developed

This class will include all commercial, urban, and residential areas. Common features within developed areas include buildings, roads, yards, and parking lots.



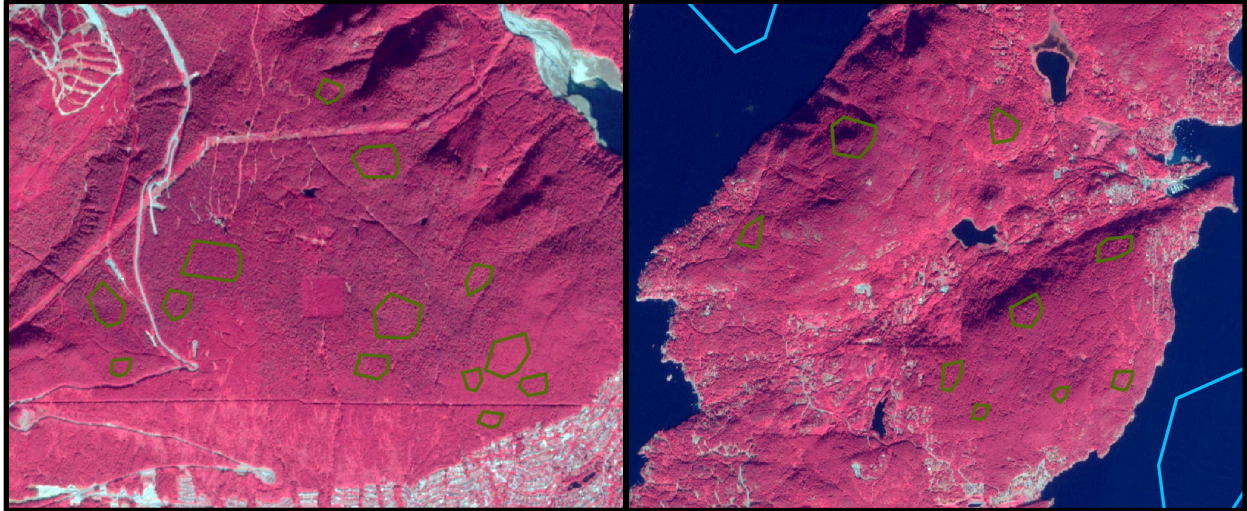
Barren

This class will include non-vegetated areas not associated with development, such as bare rock or soil. This is not a common cover type in this image. However, there are some bare rock surfaces in the mountainous area in the top part of the image.



Forest

This class will include all forested areas in the image. Do not include small stands of trees in residential areas.



Herbaceous

This class will include vegetated areas that are not trees, such as fields, pastureland, and cropland.



Water

This class will include all water features, such as the ocean, ponds, lakes, and rivers.

