GEOS451 Assignment03 — Image Classification

Patrick Harrison (000733057)
Southern Alberta Institute of Technology
GEOS451: Data Capture II
Adrian Faraguna
Novemeber 24, 2024

Submission includes 11 pages.

Contents

Introduction

The Ledum Land Trust (LLT) is a conservation organization dedicated to preserving natural areas within the City of Calgary region. Landcover such as agricultural, natural, heritage, scenic, and recreational values are of importance to LLT. As a proof of concept to LLT, an area of land near Waiparous Alberta has been selected for a landcover assessment. This report outlines a development process for a sampling strategy for Developed, Barren Ground, Tree/Shrub. Grassland and Water land cover classifications.

A critical first step in this process is gaining an understanding of the landcover types within the parcel. To achieve this, a supervised landcover classification of multispectral satellite imagery will be performed. This classification will provide valuable insights into the site's conditions and guide LLT in planning field data capture missions for their ecological assessments.

This report outlines the methodology and results of the supervised landcover classification undertaken for the Waiparous study area. The classification process utilized two algorithms, Maximum Likelihood and Support Vector Machine, on two separate satellite images of the study area, resulting in a total of four classifications. The analysis included an accuracy assessment for each classification, providing a quantitative measure of their reliability. The results of this effort include detailed landcover proportions, accuracy matrices, and recommendations for improving classification accuracy in future efforts.

Data Descriptions

Two Images were compared with the same classification parameters for both an SVM and MLE supervised classification. Both images have the same pixel size and both images are Planet Scope scenes that are ortho rectified and corrected for atmosphic distortions. This means that they are Level 2 Imagery products. A complete property comparison of the two images can be found in Table 8.

Provided Planet Scope PS2 Data

The Provided imagery is planet scope imagery taken on the 27^{th} of July. Planet scope imagery is produced by a constellation of satellites that they call Doves (Planet Labs PBC, 2020). The first given image was taken using the PS2 instrument with 4 bands. The bands included in the 4 band imagery are the standard Red, Green, Blue, and NIR found in Table 1.

The data product for the imagery is corrected for surface reflectance and is ortho rectified. This makes it suitable for a supervised classification of landcover (Planet Labs PBC, 2024).

Band	**Colour**	**PS2 Range**
Band 1	Blue	455 - 515 nm
Band 2	Green	500 - 590 nm
Band 3	Red	590 - 670 nm
Band 4	NIR	780 - 860 nm

Table 1: PS2 Instrument bands (Planet Labs PBC, 2023)

Found Planet Scope PSB.SD Data

An 8-band image was also downloaded from Planet.com for the same study area on the same day. This way, accuracy assessment points will not need to be redone between images as the landcover should not change in that short of time.

The Planet Scope Blue Super Dove (PSB.SD) image has 8 bands as seen in Table 2. This image is also orthorectified and corrected for atmospheric conditions.

Band	**Colour**	**Super Dove Range**
Band 1	Coastal Blue	431 - 452 nm
Band 2	Blue	465 - 515 nm
Band 3	Green I	513 - 549 nm
Band 4	Green	547 - 583 nm
Band 5	Yellow	600 - 620 nm
Band 6	Red	650 - 680 nm
Band 7	Red Edge	697 - 713 nm
Band 8	NIR	845 - 885 nm

Table 2: Super Dove bands (Planet Labs PBC, 2023)

Methodology

For this process, the classification wizard was used to facilitate the workflow. The entire workflow can be seen in figure 1. Starting with the raw imagery the model is prepared based off the imagery. The classification wizard was setup for a pixel based supervised classification. Training samples are chosen from the imagery and the classification is performed. Once the classified raster is generated, testing the accuracy of the model is done through creating or updating testing points and from these points a confusion matrix is generated that shows a thourough assessment of the models performance. Post processing such as a majority filter to smooth the classification results is used and zonal statistics are generated to quantify the landcover results from the model (SAIT, 2024).

Prepare the Data

Preparing the data in the model involves inspecting the type of landcover present in the image to build the model schema. The imagery gathered for this analysis was determined to be suitable as they are level 1 or level 2 satellite data products. Level 0 or raw data will not perform well with the type of classification. Level 1 or level 2 data products that are radiometrically corrected are better for this landcover assessments.

The imagery was also determined to cover 100% of the study area. Training, classification and accuracy assessments should be done on the whole image so the imagery was not clipped or masked to the study area until the reporting step.

Collect Training Data

Training samples were collected through digitization of areas in the imagery. Other imagery to delineate landcover types, only use the imagery that is being classified. It is useful to use both true-colour as well as false colour imagery such as indicies to better differentiate landcover classes. In this process an SR index was used to help identify barren ground areas as well as a modified normalized water index to find the small pockets of water in the area. Samples captured included variability within the landcover class and delineated areas are completely enclosed by the landcover by avoiding edges of the landcover.

The validation of training samples involves having a proportional representation of training data to the amount of landcover type in the image. Training samples should also be evenly distributed across the study area. This distribution across the image will better account for variability in land classes over a larger area.

Perform Supervised Classification

Once the training data is setup up in the classification wizard, the model to be used can be set up. In this case, for each image both a maximum likelihood model and support vector machine (SVM) model were used and compared.

There are no additional parameters were set for the maximum likelihood estimator model, and the SVM model was left at the default 500 samples. For the SVM, the number of samples is an important parameter that can either under train or over train your model. Too many sample per class, and the model will be over constrained to the training data and may miss classifications. Under training the model could result in missclassifications.

Generate Testing Data and Perform Accuracy Assessment

Testing points were created using a stratified random sampling method with 200 points based off the initial classification results. The accuracy assessment points for the first image had to be referenced with ground truth data manually. After the first image, the Update accuracy points tool lets us reuse the points for different classification so ground truth data did not have to be entered 800 times.

From the testing points, a confusing matrix can be generated to quantify the results of the classification. This confusion matrix gives overall accuracy, as well as class specific accuracies and the kappa coefficients for classes

After performing the accuracy assessment, starting at preparing the data again, the methodology was repeated for a different classification model, as well as both images.

Post-Processing Classified Data

To clean up a good classification, a smoothing technique was used called a majority filter. This filter looks at the neighbouring pixels in the image and reclassifies if necessary.

Use polygon to raster with multi part polygons to get a table describing the total amount of land cover in each class.

Results

The following confusion matricies show the accuracy assessments for each classification type. The best classification based on the accuracy assessment if the PSB.SD image with a SVM classification model. The κ (Kappa) value in the confusion matrix measure the difference from randomly assigning classes to the landcover. The best κ result was for the maximum likely hood estimator model for the PS2 confusion matrix (SAIT, 2024).

The accuracy results of each model are shown in Tables 4, 5, 6, and 7.

After the landcover was classed, The resulting highest accuracy raster was turned into multi-part polygons to determine the amount of area was present for each class in the study area. These results were as follows in Table 3

Developed	96.32 km^2
Barren Ground	69.31 km^2
Tree/Shrubs	3,795.24 km^2
Grassland	1,536.86 km^2
Water	0.25 km^2

Table 3: Landcover areas in square kilometers.

Class	**Developed**	**Barren Ground**	**Tree/Shrubs**	**Grassland**	**Water**	**Total**	**U. Accuracy**	**Kappa**
Developed	7	3	0	0	0	10	0.7	0
Barren Ground	0	8	1	0	1	10	0.8	0
Tree/Shrubs	0	1	147	6	0	154	0.954545	0
Grassland	0	4	7	27	0	38	0.710526	0
Water	0	1	9	0	0	10	0	0
Total	7	17	164	33	1	222	0	0
P. Accuracy	1	0.470588	0.896341	0.818182	0	-	0.851351	0
Kappa	0	0	0	0	0	-	0	0.674747

Table 4: PS2 SVM Confusion matrix.

Class	**Developed**	**Barren Ground**	**Tree/Shrubs**	**Grassland**	**Water**	**Total**	**U. Accuracy**	**Kappa**
Developed	7	3	0	0	1	11	0.636364	1
Barren Ground	0	9	3	1	0	13	0.692308	0
Tree/Shrubs	0	0	153	7	0	160	0.95625	0
Grassland	0	4	8	25	0	37	0.675676	0
Water	0	1	0	0	0	1	0	0
Total	7	17	164	33	1	222	0	0
P. Accuracy	1	0.529412	0.932927	0.757576	0	-	0.873874	0
Kappa	0	0	0	0	0	-	0	0.711206

Table 5: PS2 MLE Confusion matrix.

Class	Developed	Barren Ground	Tree/Shrubs	Grassland	Water	Total	U. Accuracy	Kappa
Developed	7	3	0	0	0	10	0.7	0
Barren Ground	0	7	0	1	0	8	0.875	0
Tree/Shrubs	0	1	156	9	0	166	0.939759	0
Grassland	0	5	5	23	0	33	0.69697	0
Water	0	1	3	0	1	5	0.2	0
Total	7	17	164	33	1	222	0	0
P. Accuracy	1	0.411765	0.95122	0.69697	1	-	0.873874	0
Kappa	0	0	0	0	0	-	0	0.700578

Table 6: PSB.SD SVM Confusion matrix.

Class	Developed	Barren Ground	Tree/Shrubs	Grassland	Water	Total	U. Accuracy	Kappa
Developed	7	5	1	0	1	14	0.5	0
Barren Ground	0	7	0	1	0	8	0.875	0
Tree/Shrubs	0	0	158	11	0	169	0.934911	0
Grassland	0	5	5	21	0	31	0.677419	0
Water	0	0	0	0	0	0	0	0
Total	7	17	164	33	1	222	0	0
P. Accuracy	1	0.411765	0.963415	0.636364	0	-	0.869369	0
Kappa	0	0	0	0	0	-	0	0.683029

Table 7: PSB.SD SVM Confusion matrix.

Discussion

This process only used 200 accuracy assessment points making the confusion matricies not very reliable to how the model actually performed. The water classification was difficult for the model and the assessment as there was not alot of training data for water in the imagery, and what was there spectrally looked very similar to shadowed trees.

To improve this model, more water training data would be nice to have and if given more time, more accuracy assessment points could be used to get a better grasp on how each individual class performed over the image.

Improvements to the model would include having a SWIR band (this may pick up water better as many water indicies use it) (Henrich, 2024). This would also be handy to have for creating the training data and being able to distinguish water from other landcover types.

Targeted accuracy assessment points could be used to improve the accuracy assessment by targeting trouble regions in the image. This would better represent how each landcover class performed.

Changing the selection of training data may help the model, or it may produce more inaccurate results. Using indicies or a modified tassel cap transformation (fixed PCA) for planet scope data would be a good option to choosing better training data.

Classification results can also be improved with using a majority filter across the image and will smooth the image and classifications. This may help with really noisy classification results.

Conclusion

Planet Imagery can be very useful in combination with supervised classification for delineating landcover types. Overall, the 8-band imagery did not significantly outperform the 4band imagery, however it did do better with the SVM than the MLE.

References

Henrich, V. (2024). IDB - List of available Indices. https://www.indexdatabase.de/db/i.php

Planet Labs PBC. (2020). Planet application program interface: In space for life on earth. *Planet*. https://api.planet.com

Planet Labs PBC. (2023). Planet imagery product specifications. *Planet*. https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf

Planet Labs PBC. (2024). *Planet*. https://developers.planet.com/docs/data/planetscope/

SAIT. (2024). Module06 - image processing. In *GEOS 451: Data Capture II*. [PDF document]. D2L/Brightspace. https://learn.sait.ca

Appendix A

Parameter	Given PS2 4 Band Imagery	Found 8 Band Imagery
Acquired	2020-07-27T18:22:30.014013Z	2020-07-27T17:56:54.587402Z
Anomalous Pixels	0.01	0
Clear Confidence Percent	99	100
Clear Percent	100	100
Cloud Cover	0	0.0
Cloud Percent	0	0
Ground Control	True	True
GSD (Ground Sample Distance)	3.9	3.9
Heavy Haze Percent	0	0
Instrument	PS2	PSB.SD
Item Type	PSScene4Band	PSScene
Light Haze Percent	0	0
Pixel Resolution	3	3
Provider	PlanetScope	PlanetScope
Published	2020-07-27T22:01:24Z	2021-03-17T01:22:37Z
Publishing Stage	Finalized	Finalized
Quality Category	Standard	Standard
Satellite ID	1004	222b
Shadow Percent	0	0
Snow/Ice Percent	0	0
Strip ID	3604072	3604017
Sun Azimuth	144.5	135.4
Sun Elevation	53.5	50.9
Updated	2020-07-28T04:57:46Z	2023-05-06T23:31:04Z
View Angle	3	5
Visible Confidence Percent	99	100
Visible Percent	100	100

Table 8: Comparison of 4 band and 8 band image properties for Planet Scope Scenes.

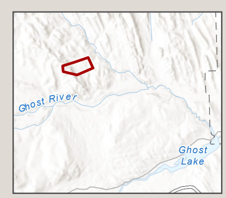
LANDCOVER CLASSIFICATION FOR THE WAIPAROUS AREA (2020)

GEOS 451 | Assignment 03

Results of the a pixel based, supervised, Support Vector Machine (SVM) landcover classification for Planet Scope Scene data. The Planet Scope Scene data here is the eight band analytical product corrected for surface reflectance. The resulting classification has an overall accuracy of 87.4% with a kappa value of 70.1%.

The water classification (not shown in actual study area) was determined to be 100% however by visual inspection, this is not the case. The 200 accuracy points were not dense enough to get a good metric for the model performance.

Related to the water classification, there were many issues differentiating between shadowed trees and water. Another common issue with this classification is birght barren ground with developed areas.



Confusion Matrix for 200 Stratified Random Assessment Points

Landcover Class	Developed	Barren Ground	Tree / Shrubs	Grassland	Water	Total	U Accuracy	Kappa
Developed	7	3	0	0	0	10	0.700	0.000
Barren Ground	0	7	0	1	0	8	0.875	0.000
Tree / Shrubs	0	1	156	9	0	166	0.940	0.000
Grassland	0	5	5	23	0	33	0.697	0.000
Water	0	1	3	0	1	5	0.200	0.000
Total	7	17	164	33	1	222	0.000	0.000
P_Accuracy	1	0.411765	0.95122	0.69697	1	0	0.874	0.000
Kappa	0	0	0	0	0	0	0.000	0.701

0 0.5 1

Kilometers

Scale: 1:12,000

Projection: NAD83 UTM Zone 11N

N

Data Source:
Planet Labs PBC, SAIT

Cartography: Patrick Harrison (733057)

Date: 2024/11/24

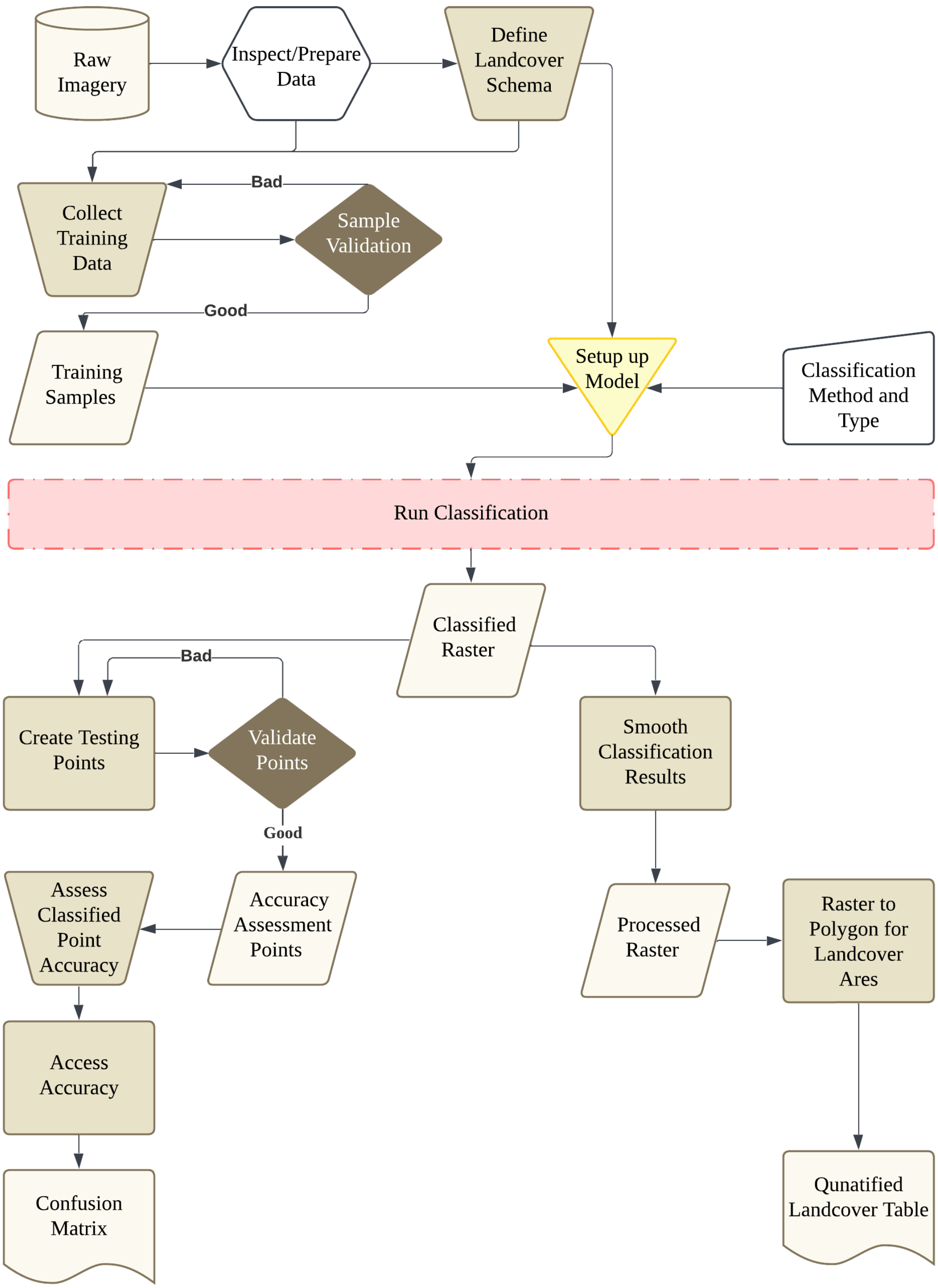


Figure 1: Supervised Pixel based classification workflow.