

AUTOMATIC VEGETATION CLASSIFICATION USING NDVI

KOLTOUCH ANANTAKARN

MARWIN SUWANPRAPA



A THESIS REPORT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE BACHELOR OF

ENGINEERING PROGRAM IN CIVIL ENGINEERING

INTERNATION PROGRAM

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2021

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use

AUTOMATIC VEGETATION CLASSIFICATION USING NDVI

KOLTOUCH ANANTAKARN

MARWIN SUWANPRAPA



A THESIS REPORT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE BACHELOR OF

ENGINEERING PROGRAM IN CIVIL ENGINEERING

INTERNATION PROGRAM

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2021

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use



This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use

Thesis Certification
International College
King Mongkut's Institute of Technology Ladkrabang

Thesis Title Automatic Vegetation Classification using NDVI

Student Mr Koltouch Anantakarn
Mr Marwin Suwanprapa


Student ID 60011225
60011230

Degree Bachelor of Engineering

Program Civil Engineering (International Program)

Thesis Advisor Dr. Chisaphat Supunyachotsakul

Thesis Co-Advisor Dr. Nobphadon Suksangpanya

EXAMINERS	SIGNATURES
Dr. Chisaphat Supunyachotsakul	
Dr. Nobphadon Suksangpanya	
Prof. Uba Sirikaew	
Prof. Pinit Tanachaichoksirikun	

Date: May 31st, 2021

Time: 13.00-15.00

Place: King Mongkut's Institute of technology Ladkrabang



(Asst. Prof. Dr. Arthit Petchsasithon)

Dean of the Department of Civil Engineering

May 31st, 2021

Thesis Title	Automatic Vegetation Classification using NDVI
Student	Mr Koltouch Anantakarn Mr Marwin Suwanprapa
Student ID	60011225 60011230
Degree	Bachelor of Engineering
Program	Civil Engineering (International Program)
Thesis Advisor	Dr. Chisaphat Supunyachotsakul
Thesis Co-Advisor	Dr. Nobphadon Suksangpanya

ABSTRACT

Due to climate change and deforest issues green area has become one of the most important factors to prevent a damage. In Thailand and other country around the world try to increase a green area not only to fight against climate and deforest issues, but also to improve air quality. To observe a green area in a large-scale using satellite image are eligible. Satellite images offer a useful information to produce various outputs. To survey for a green area NDVI are beneficial since satellite image are in charge. NDVI is an indicator that can be used to analyze remote sensing measurements, assessing whether the target being observed contains live green vegetation. Although NDVI are good indicators to observe a green area, however NDVI still cannot classified a green area without thresholds. To find a proper threshold for NDVI classification we have been through area type of area experiment and temporal area experiment. Both experiments aim to observe a proper threshold that satisfied a high prediction accuracy. Finding threshold for NDVI are required ground truth which is a trustable information of a green area and NDVI image. To provide a ground truth we used digitizing method to obtain the information. Type of area experiment is to use different type of area as a variant and for temporal area we use time as a variant. Temporal area experiment is to

observe the same are with different timing to compare the results. As the results we found that at threshold around 0.325 to 0.367 are suited to observe a green area.

Keywords: NDVI; Satellite image; Remote Sensing; Digitizing; Ground truth; Threshold.



ACKNOWLEDGEMENT

Without the contribution of many people, this thesis would not have been existed. It owes the existence to the supports and inspirations from a lot of people.

To our thesis advisor Dr. Chisaphat Supunyachotsakul of school of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang and our thesis co-advisor Dr. Nobphadon Suksangpanya of School of Geoinformatics, Institute of Science, Suranaree University we would like to express our deepest gratitude for the encouragement and supervision through all obstacles and challenges since the beginning until the end of our study.

We also want to express my gratitude to all lecturers for your support and guidance to ours for the whole two years. Also, we would like to thank all of our friends who always be there to support and motivate us as always. Moreover, we also would love to express our gratitude to all respondents who contribute their information and time on this study. And we do believe the study could not been done without their input.

Finally, we must express our very greatest gratitude to our parents and all relatives for providing us with unfailing support and continuous motivation throughout our years of study. This accomplishment would not have been possible without them.

Koltouch Anantakarn

Marwin Suwanprapa

TABLE OF CONTENTS

CHAPTER	PAGE
ABSTRACT.....	I
ACKNOWLEDGEMENT.....	III
TABLE OF CONTENTS.....	IV
LIST OF TABLES.....	VII
LIST OF FIGURES.....	VIII
CHAPTER 1 INTRODUCTION.....	1
1.1 Research Background.....	1
1.2 Objectives.....	2
1.3 Research Plans.....	3
CHAPTER2 LITERATURE REVIEWS.....	5
2.1 Temporal optimization of image acquisition for land cover classification with Random Forest and MODIS time-series.....	5
2.2 Very High-Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting.....	7
2.3 A Python-Based Open-Source System for Geographic Object-Based Image Analysis (GEOBIA) Utilizing Raster Attribute Tables.....	9

TABLE OF CONTENTS

(Continued)

CHAPTER	PAGE
2.4 Classification of Landsat 8 Satellite Data Using NDVI Thresholds.....	11
CHAPTER 3 BACKGROUND THEORIES.....	12
3.1 NDVI.....	12
3.2 Landsat Collection 2.....	15
3.3 Comparison Landsat 8 collection 1 and 2.....	22
3.4 Confusion Matrix.....	23
CHAPTER 4 RESEARCH METHODOLOGY.....	26
Overview of Research Methodology.....	26
4.1 Finding NDVI.....	27
4.2 Converting to RGB.....	28
4.3 Digitizing.....	29
4.4 Finding proper thresholds.....	33
4.4.1 Confusion matrix.....	33
4.4.2 Finding Accuracy.....	36

TABLE OF CONTENTS

(Continued)

CHAPTER	PAGE
CHAPTER 5 RESULTS AND DISCUSSIONS.....	50
5.1 Results.....	50
5.1.1 Three types of area result.....	50
5.1.2 Temporal area experiment.....	52
5.1.3 Compare ground truth results.....	54
5.2 Discussions.....	54
CHAPTER 6 CONCLUSIONS AND FUTURE WORK.....	55
6.1 Conclusions.....	55
6.2 Future work.....	55
REFERENCES.....	56
AUTHORS BIOGRAPHY.....	57

LIST OF TABLES

TABLE	PAGE
Table 1.1 2020 Research plan.....	4
Table 2.1 Result in Dakar, Senegal.....	8
Table 2.2 Ouagadougou, Burkina Faso.....	8
Table 3.1 Table of bands in Landsat 8-9 and Thermal sensors.....	14
Table 3.2 Collection 2 Surface reflectance enhancement.....	20
Table 3.3 Collection 2 Surface temperature enhancement.....	21
Table 3.4 Summary between Landsat collection 1 vs Landsat collection 2.....	22
Table 3.5 Confusion Matrix.....	23
Table 4.1 Confusion Matrix.....	33
Table 4.2 Table of maximum accuracy and thresholds form temporal experiment.....	49
Table 5.1 Results of Area types experiment.....	51
Table 5.2 Results of temporal area experiment.....	52

LIST OF FIGURES

FIGURES	PAGE
Figure 2.1 Seasonal mean land cover class NDVI curves with one standard deviation above and below the mean shown in the shaded region.....6	6
Figure 2.2 Seasonal mean land cover class EVI curves with one standard deviation above and below the mean shown in the shaded region.....6	6
Figure 2.3 Seasonal mean Grassland classes with one standard deviation above and below the mean shown in the shaded region – a) NDVI, b) EVI.....7	7
Figure 2.4 The graph that show mean Overall accuracy all classifiers.....9	9
Figure 2.5 Raster clumps used to store an object representation.....10	10
Figure 2.6 Flow chart of segmentation algorithm.....10	10
Figure 2.7 Overview of NDVI Thresholds Studies.....11	11
Figure 3.1 The picture show healthy vegetation reflects more NIR and Green light....12	12
Figure 3.2 Day Transition Fixed.....16	16
Figure 3.3 The images display the estimated change of Landsat 8 Thermal band calibration update.....17	17

LIST OF FIGURES

(Continued)

FIGURES	PAGE
Figure 3.4 Left: Sample of a TIRS Band 10 image with radiometric striping when processed into Collection 1 and Right: The reduction of striping in Collection 2.....	18
Figure 3.5 Landsat 8 OLI images of collection 1 on the left and collection 2 on the right.....	19
Figure 3.6 Landsat collection 2 examples.....	19
Figure 4.1 Research methodology overview.....	26
Figure 4.2 NDVI image converting form LANDSAT 8 collection2.....	28
Figure 4.3 RGB image.....	29
Figure 4.4 Clipped RGB imagery.....	30
Figure 4.5 Digitizing process.....	30
Figure 4.6 Raw ground truth form digitizing process.....	31
Figure 4.7 Ground truth 1 (Koltouch).....	32
Figure 4.8 Ground truth 1 (Marwin).....	32
Figure 4.9 Area overlap ground truth.....	32
Figure 4.10 Example of classified NDVI raster.....	34
Figure 4.11 Examples of NDVI threshold classifier.....	45

LIST OF FIGURES

(Continued)

FIGURES	PAGE
Figure 4.12 plotting (Threshold, Accuracy%)	37
Figure 4.13 Overview of area, each square represented different types of terrain.....	38
Figure 4.14 Urban area.....	39
Figure 4.15 Rural area.....	40
Figure 4.16 Forest area.....	40
Figure 4.17 Plot of accuracy and thresholds of Urban area scene 1.....	41
Figure 4.18 Plot of accuracy and thresholds of Urban area scene 2.....	41
Figure 4.19 Plot of accuracy and thresholds of Rural area scene 1.....	42
Figure 4.20 Plot of accuracy and thresholds of Rural area scene 2.....	42
Figure 4.21 Plot of accuracy and thresholds of Forest area scene 1.....	43
Figure 4.22 Plot of accuracy and thresholds of Forest area scene 2.....	43
Figure 4.23 Green area Prediction with maximum accuracy of urban area scene 1.....	44
Figure 4.24 Green area Prediction with maximum accuracy of rural area scene 1.....	44
Figure 4.25 Green area Prediction with maximum accuracy of forest area scene 1.....	45
Figure 4.26 Green area Prediction with maximum accuracy of urban area scene 2.....	45

LIST OF FIGURES

(Continued)

FIGURES	PAGE
Figure 4.27 Green area Prediction with maximum accuracy of rural area scene 2.....	46
Figure 4.28 Green area Prediction with maximum accuracy of forest area scene 2.....	46
Figure 4.29 RGB image converted form satellite image taken in 19/2/2020.....	47
Figure 4.30 RGB image converted form satellite image taken in 29/8/2020.....	48
Figure 4.31 RGB image converted form satellite image taken in 17/11/2020.....	48
Figure 5.1 Green area prediction at 0.367 threshold compared to RGB.....	51
Figure 5.2 Green area prediction at 0.325 threshold compared to RGB (Taken at 19/02/2020)	52
Figure 5.3 Green area prediction at 0.325 threshold compared to RGB (Taken at 29/08/2020)	53
Figure 5.4 Green area prediction at 0.325 threshold compared to RGB (Taken at 17/11/2020)	53

CHAPTER 1

INTRODUCTION

1.1 Research Background

In 2020, we are facing to the world biggest issues, the global warming. We often call the result “global warming”, but it is causing a set of changes to the Earth's climate, or long-term weather patterns, that varies from place to place. While many people think of global warming and climate change as synonyms.

Climate change encompasses not only rising average temperatures but also extreme weather events, shifting wildlife populations and habitats, rising seas, and a range of other impacts. All of those changes are emerging as humans continue to add heat-trapping greenhouse gases to the atmosphere, changing the rhythms of climate that all living things have come to rely on.

The global warming main cause is come from the greenhouse effect. The "greenhouse effect" is the warming that happens when certain gases in Earth's atmosphere trap heat. These gases let in light but keep heat from escaping, like the glass walls of a greenhouse, hence the name. In the past, this greenhouse gas is balance to natural absorbing but after people known how to use conventional energy such as fuels, they emitted the greenhouse gases more than natural can absorb. It means that this gas will be in the atmosphere for a long time until natural will absorb all of it.

The most common and most talked about greenhouse gases is CO₂ or carbon dioxide. In fact, because it is so common, scientists use it as the benchmark or measure of things that warm the atmosphere. CO₂ or carbon dioxide is produced any time

something is burned. It is the most common greenhouse gases, constituting by some measures almost 55% of total long-term greenhouse gases.

The impact of the global warming is now happened around the world. The impact that are obviously see is a risen of Earth's temperature. Not only that, melting glaciers, flooding, droughts and even extinction of an animal are the impacts from the global warming.

In conclusion, we are concern that a deforestation is one of the main issues that making natural cannot absorb the greenhouse gases quickly enough to make the greenhouse gases gone from the atmosphere. So, our project is aimed for writing an algorithm to find the specific detail of a green area from the satellite image that cannot know by using only your eyes and make the program can easily use widely.

1.2 Objectives

1. To study the beneficial of NDVI for finding a green area in a large-scale work.
2. To find the proper threshold of NDVI to observe a green area.
3. To find that our threshold is accurate enough to use for finding a green area

1.3 Research Plans

Research plan was used to set up and plan an overview of the project. As shown in **Table 1.1** research plan shown timing when each step has done. In research plan each step was categorize as information, program, coding, and correction. Here is the definition and working scope in each step.

1. Satellite imagery: In this step we search for an information about satellite image and how it is work.
2. NDVI: This step we studied normalized difference vegetation index (NDVI) and how NDVI can get a vegetation indexes using satellite image.
3. Literature review: Is to search and evaluation of the available literature that relevant to our project's issues.
4. MATLAB: Studied MATLAB program and work with some information mostly in matrix.
5. Python modules: Learn how to use python3 and their modules.
6. Python workshop: Starting to use python as a tool to work with satellite images.
7. NDVI: Finding NDVI raster form satellite images for classification methods.
8. Classify green area: Classify a vegetation form a particular area to find a proper threshold.
9. Digitizing: Observing green area manually using ArcGIS program to use as a reference for classification models.
10. Confusion matrix: Studied about confusion to apply to project's models to find an accuracy of green area predictions.

11. Area types: Studies about thresholds that satisfied high accuracy for different types of areas. The areas include urban area, rural area, and forest area.

12. Landsat collection 1,2: Studied and compared Landsat collections 1,2.

Table 1.1 Research plan

Research plan	2020						2021					
	6	7	8	9	10	11	12	1	2	3	4	5
Information												
Satellite Imagery	■											
NDVI	■											
Literature review		■	■	■								
Program												
MATLAB		■										
Python modules			■									
Python workshop			■	■								
Coding												
Rasterio				■	■							
NDVI					■	■						
Classify green area						■	■	■				
Correction												
Digizing								■	■	■	■	
Confusion matrix									■	■	■	■
Area types										■	■	
Temporal area												■
Landsat collection 1,2											■	■

CHAPTER 2

LITERATURE REVIEW

2.1 Temporal optimization of image acquisition for land cover classification with Random Forest and MODIS time-series.

The report is about comparing NDVI and MODIS EVI (Enhance Vegetation Index) that more accurately and can show sharper growing season peaks and greater sensitivity by timeseries (such as Hodrick-Presscott or Gaussian filters), different smoothing techniques, ranging from Fourier based filters.

They extracted NDVI and EVI values from MOD13Q1 and arranged to create a time series stack, use the filters to eliminate the error values and Random Forest to cut out-of-bag error. After that, plot by using Fourier based filter.

The result shows in the **Figure of 2.1, 2.2 and 2.3** below that it would be great to use EVI value to classify between Grassland and forest and use NDVI value to classify Settlement and Peatland because it is less photosynthetically active classes.

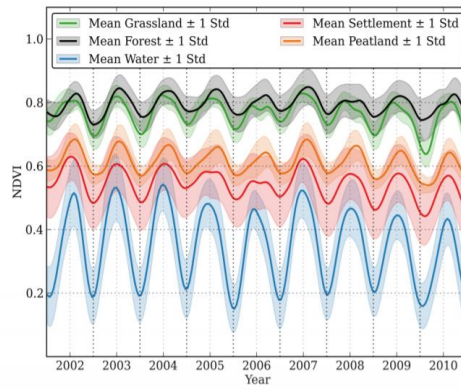


Figure 2.1 Seasonal mean land cover class NDVI curves with one standard deviation above and below the mean shown in the shaded region

(Source: Nitze and Cawkwell, 2015)

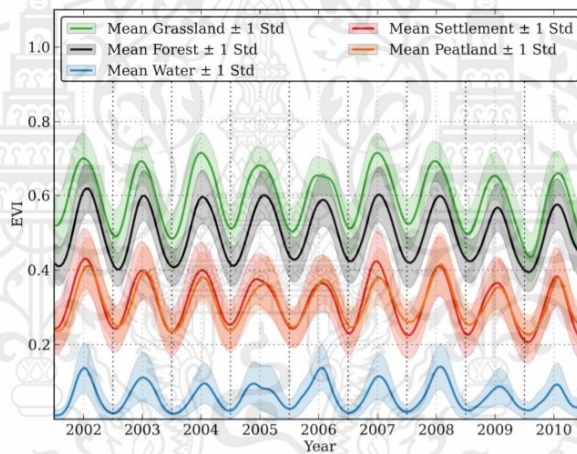


Figure 2.2 Seasonal mean land cover class EVI curves with one standard deviation above and below the mean shown in the shaded region

(Source: Nitze and Cawkwell, 2015)

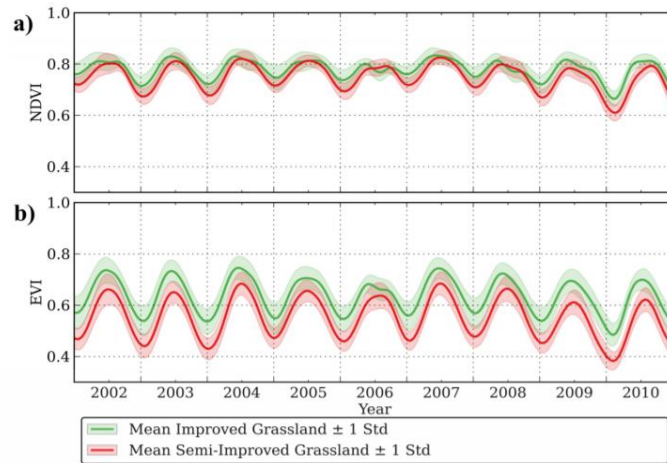


Figure 2.3 Seasonal mean Grassland classes with one standard deviation above and below the mean shown in the shaded region – a) NDVI, b) EVI

(Source: Nitze and Cawkwell, 2015)

2.2 Very High-Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting.

The report is about the authors want to compare a new invention, regularized implementation of GBM (Gradient Boosting Machine), called extreme gradient boosting (Xgboost) with the traditional boosting algorithms, Random Forest and Support Vector Machine to find the most accurate and less time processing.

They got the picture from 3 different places and segmented the data by using Grass GIS. First and Second place using random and stratified random sampling with strata defined using OpenStreetMap and make 3 training sets of 20, 40 and 60. The last place, they used the sensitivity to sample size in more depth (between 50 and 400 objects per class) while we also used ten randomly drawn distributions of training objects for each sample size category. After that put the data of 3 different places to all boosting algorithms and record the result.

From the **Table 2.1, 2.2** and **Figure 2.4** below, Xgboost performed very good overall accuracy when using medium and large sample size, but it cost increased computational time. The benefits of CFS (Correlation Feature Selection) as FS (Feature Selection) method for reducing dimensionality are more beneficial for small sample sizes.

Table 2.1 Result in Dakar, Senegal

Training sample	SVM	SVM _{CFS}	RF	RF _{CFS}	Xgb	Xgb _{CFS}
N ₂₀	74.65	75.85	77.12	77.67	74.81	75.56
N ₄₀	77.12	77.31	78.33	79.4	80.09	79.25
N ₆₀	77.88	78.07	80.33	80.88	82.1	81.78
Time (s)	0.25	0.18	1.51	1.01	3.62	2.21

(Source: Geoganos et al., 2018)

Table 2.2 Ouagadougou, Burkina Faso

Training sample	SVM	SVM _{CFS}	RF	RF _{CFS}	Xgb	Xgb _{CFS}
N ₂₀	84.53	83.53	82.57	82.55	83.42	84.23
N ₄₀	86.74	85.64	85.83	86.19	87.21	87.30
N ₆₀	88.39	88.02	86.78	86.69	88.44	88.10
Time (s)	0.29	0.21	1.75	0.82	4.15	1.93

(Source: Geoganos et al., 2018)

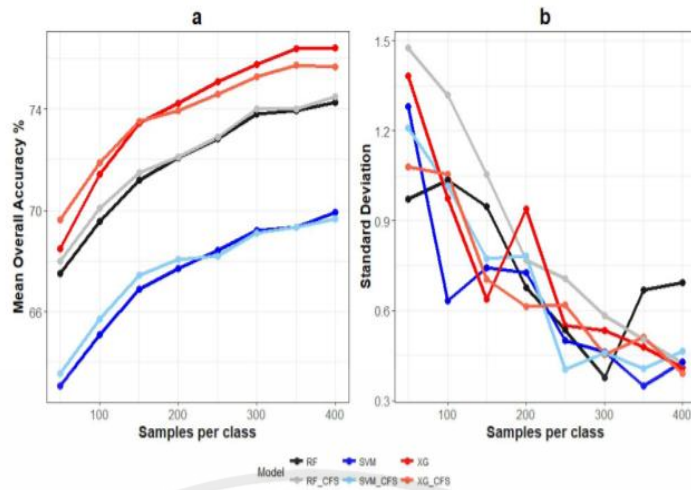


Figure 2.4 The graph that show mean Overall accuracy all classifiers
(Source: Geoganos et al., 2018)

2.3 A Python-Based Open-Source System for Geographic Object-Based Image Analysis (GEOBIA) Utilizing Raster Attribute Tables

This report performs a formula system for Geographic Object-Based Image Analysis (GEOBIA), using entirely open-source software. System is presented as raster clumps and storing attributes as a raster attribute table (**Figure 2.5**). The system is particularly suited to very high-resolution multispectral imagery but lower resolution data such as Landsat can also be used, but with fewer classes derived.

They used raster clumps to predict a type of area and classified. The research methodology also useful as you can see in (**Figure 2.6**). Segmentation using implementation of K-means clustering to generate seeds for the segmentation where, following the assignment of the pixels to the associated cluster center, the clumps are iteratively eliminated if they are below the minimum mapping unit threshold to the neighboring clump that is closest in “color”. Following elimination, the final clumps are relabeled to ensure that they are consecutively numbered, where a value of zero

defines no data regions. The algorithm is repeatable, always producing the same result given the same input data and parameters.

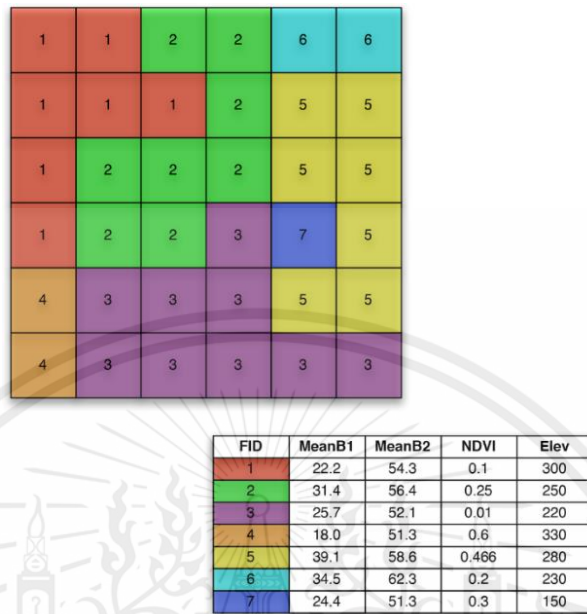


Figure 2.5 Raster clumps used to store an object representation

(Source: Clewley et al., 2014)

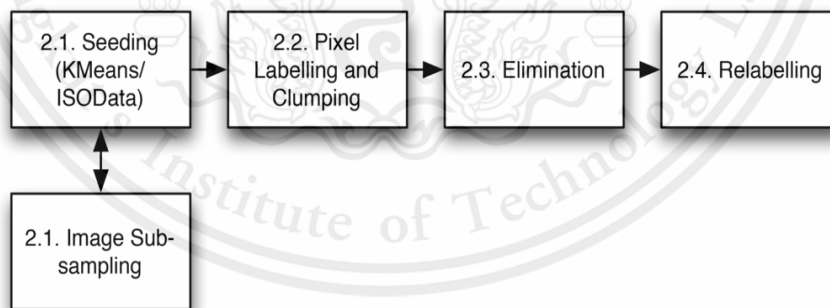


Figure 2.6 Flow chart of segmentation algorithm

(Source: Clewley et al., 2014)

2.4 Classification of Landsat 8 Satellite Data Using NDVI Thresholds

This study goals are to classify Landsat 8 satellite data using NDVI thresholds. Initially, visible and near infrared bands of Landsat 8 satellite were used to derive Normalized Different Vegetation Index (NDVI) image. Vegetation, non-vegetation and water areas were then analyzed where thresholds for separating them are carefully determined with the aid of ground truth information of the study area. Their rescare plan are useful to our project in accuracy assessment section shown in **Figure 2.7**.

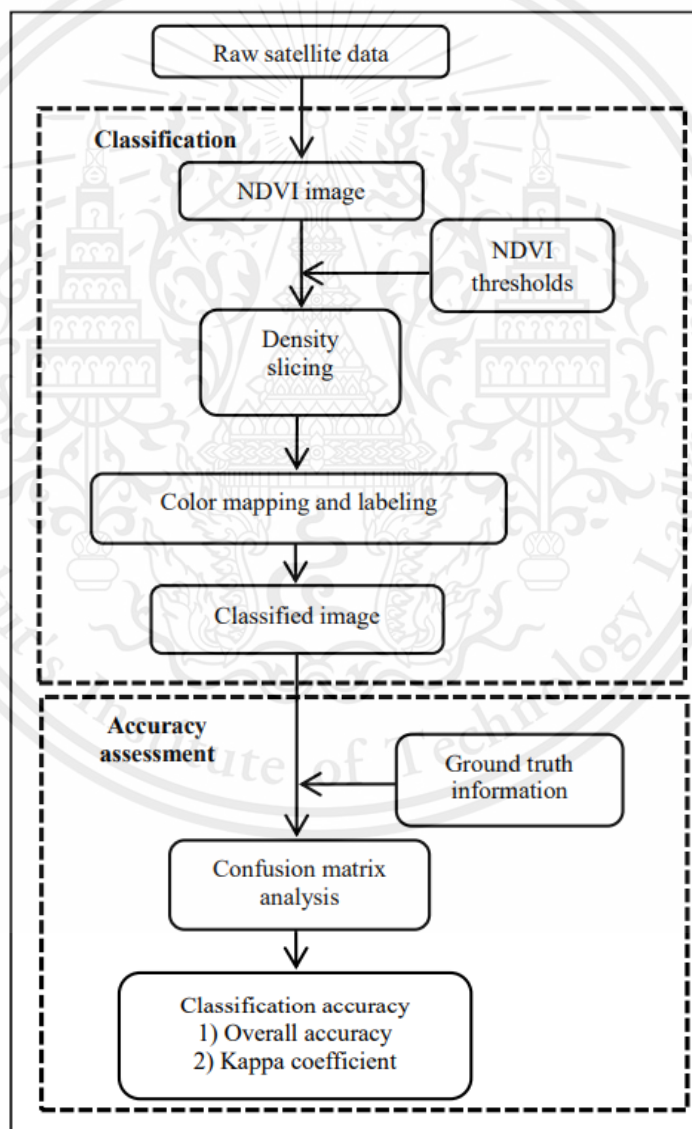


Figure 2.7 Overview of NDVI Threshold Studies

(Source: Taufik et al., 2016)

This material is reserved for educational use only, not allowed for commercial use.

CHAPTER 3 BACKGROUND THEORIES

3.1 NDVI

NDVI or also known as Normalized Difference Vegetation Index is a vegetation quantifier by measuring the difference between near- infrared which vegetation strongly reflects and red light which vegetation absorbs (**Figure 3.1**). NDVI is calculated in accordance with the formula:

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED}) \quad (3.1)$$

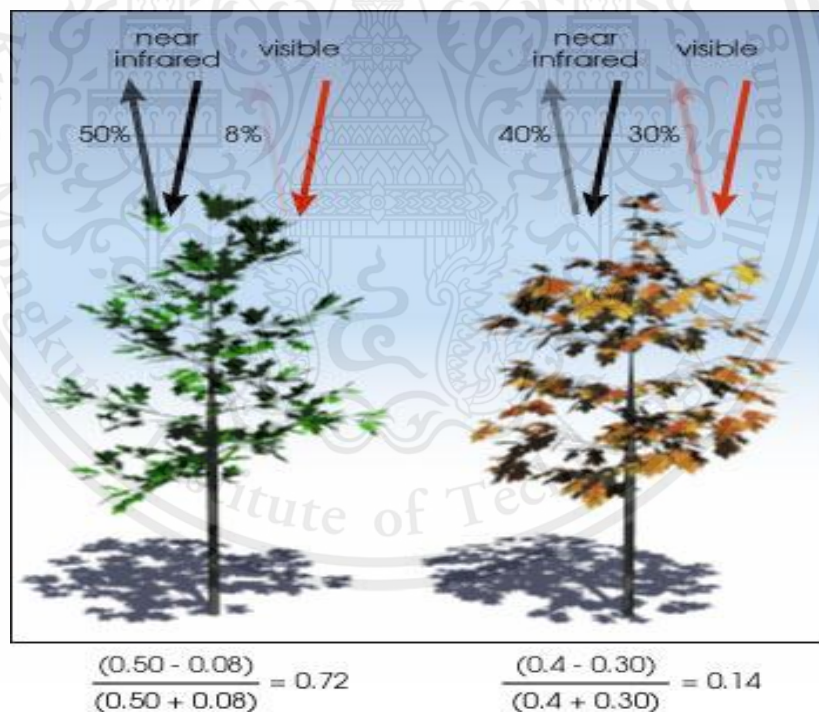


Figure 3.1 The picture show that healthy vegetation reflects more NIR and Green light
(Source: <https://gisgeography.com/ndvi-normalized-difference-vegetation-index/>)

Healthy vegetation reflects more NIR and green light (that why we see vegetation as green) and absorb more red and blue light. NDVI always range from -1 to 1 but there is not a distinct boundary for each type of land cover.

For example, if NDVI is negative it is likely to be water, if it close to +1 it is likely to be dense green leaves and if $NDVI \approx 0$ it could be urbanized area. In our case we calculate NDVI using imagery from Landsat.

In Landsat 4-7,

$$NDVI = \frac{Band4 - Band3}{Band4 + Band3} \quad (3.2)$$

In Landsat 8,

$$NDVI = \frac{Band5 - Band4}{Band5 + Band4} \quad (3.3)$$

And you can see all the band's information of Landsat 8-9 to use in the formular in the **Table 3.1**.

Table 3.1 Table of bands in Landsat 8-9 and Thermal sensors

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43-0.45	30
Band 2 - Blue	0.45-0.51	30
Band 3 - Green	0.53-0.59	30
Band 4 - Red	0.64-0.67	30
Band 5 - Near Infrared (NIR)	0.85-0.88	30
Band 6 - SWIR 1	1.57-1.65	30
Band 7 - SWIR 2	2.11-2.29	30
Band 8 - Panchromatic	0.50-0.68	15
Band 9 - Cirrus	1.36-1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.60-11.19	100
Band 11 Thermal Infrared (TIRS) 2	11.50-12.51	100

(Source: <https://gisgeography.com/ndvi-normalized-difference-vegetation-index/>)

3.2 Landsat Collection 2

3.2.1 Landsat collection 2 Level-1

Level-1 Data improvements

Data enhancements at the first level Landsat Collection 1 Level-1 scenes with a Digital Elevation Model (DEM) value greater than the elevation source's geographical area (-57.0 degrees latitude) are processed exclusively to a Level-1 Systematic Terrain (L1GT) product. Removing this constraint allows for more extreme latitude acquisitions to process to a Level-1 Terrain Precision (L1TP) product. For example, this is most pronounced at the southern tip of South America (path 227/row 99).

- **Landsat 8 Instrument Alignment:** The Thermal Infrared Sensor (TIRS) Scene Select Mirror (SSM) anomaly on Landsat 8 occurred in December 2014. The SSM current encoder was turned off after reaching the "yellow limit (-12 V) thresholds". As a result, the Anomaly Review Board (ARB) was created as a new TIRS with a redundant set of electronics that can be swapped between side A (primary) and side B (redundant/backup) if the primary electronics fail.
- **Remove GCPs Constraint in Level-1 Processing:** In Landsat 4-5 TM and Landsat 7 ETM+ Level-1 processing, a limitation of 250 GCPs is used in Collection 1. The number of acquisitions that can be processed into a Level-1 Terrain Precision (L1TP) product is limited by this GCP restriction. Together with the European Space Agency's Copernicus Sentinel-2 Global Reference Image (GRI) harmonization, removing this constraint allows more Landsat 4-7 scenes to process to a L1TP product.

Geometric Accuracy Improvements

- **Ground Control Points Phase 4:** For Collection 2 data, the Landsat 8 OLI GCPs were re-baselined to the Copernicus Sentinel-2 GRI of the European Space Agency. This entailed updating over 5.1 million GCPs and extracting 2.5 million new Landsat 8 GCPs around the world.
- **Increase DEM Scene Framing Extent:** In some Collection 1 scenes, there was a duplicated "wedge" of data in the scene's extreme southwest corner. This mostly happened in Band 1 since a DEM tile did not cover enough of the scene. Scenes prone to this issue are caught by all elevation tiles, and the scene is exactly terrain adjusted, thanks to an increase in the DEM framing extent.

- **Day Transition Fix Error:** This problem caused some of the time-dependent precision correction parameters to be wrong, and in certain situations, the error was big enough to produce considerable warping in the Terrain Precision (L1TP) image or even cause the image to fail processing entirely. As this error only affected scenes where the day of acquisition changed within the start and end times of the scene, there are only a few scenes per day that were impacted. The **Figure 3.2** below show how the Collection 2-processed scene appears after the fix was implemented. On the left is a warping that can be clearly seen in the Collection 1-processed scene and right illustrates how the Collection 2-processed scene appears after the fix was implemented.

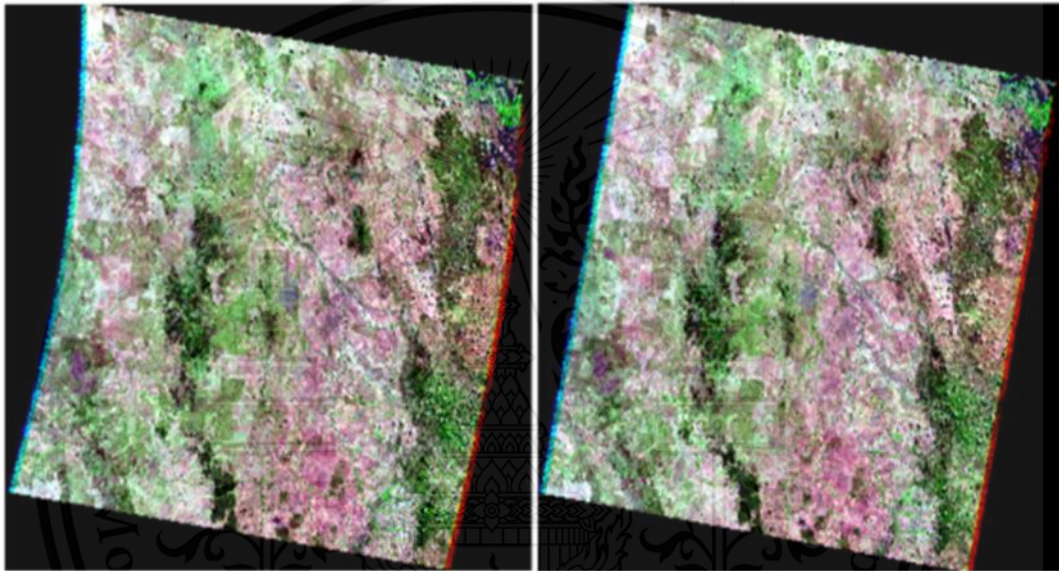


Figure 3.2 Day Transition Fixed

(Source: <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-day-transition-error-fix>)

Radiometric Accuracy Improvements

- **Landsat 8 TIRS Absolute Radiometric Calibration Update:** The radiometric calibration errors were driven by stray light. These vicarious calibrations used instrumented buoys in various water bodies and covered temperatures from ~275 K to ~300 K. After assembling the lifetime of vicarious calibration results, the calibration team found that the absolute calibration for both bands needed to be revised after compiling the lifespan of vicarious calibration findings. The TIRS instrument was calibrated on both electronics sides prior to launch. These pre-launch calibration parameters have been used to convert the raw image data to radiance since the stray light correction was implemented (the initial correction for an average error introduced by stray light that was applied in the processing system in 2014 was removed with the Collection 1 implementation

This material is reserved for educational use only, not allowed for commercial use.

of the comprehensive stray light correction (Barsi, 2014). The calibration parameters are changed in this update to account for the faults discovered by the vicarious calibration. A linear gain and bias adjustment are used in the calibration update. As shown in the **Figure 3.3**, plot the errors in temperature and radiance per-band. The calibration parameters have been adjusted to remove these errors, and Landsat 8 TIRS data processed into Collection 2 (estimated to begin in March of 2020) will contain updated radiometric calibration. For both operational modes of the sensor, these images show the estimated change in response as a result of the calibration update for each Landsat 8 Thermal band for a range of top-of-atmosphere radiances and comparable temperatures. On Side-B, the change is much more signal-dependent. Positive change indicates that the Collection 1 response is too hot, negative change indicates that the Collection 1 response is too cold.

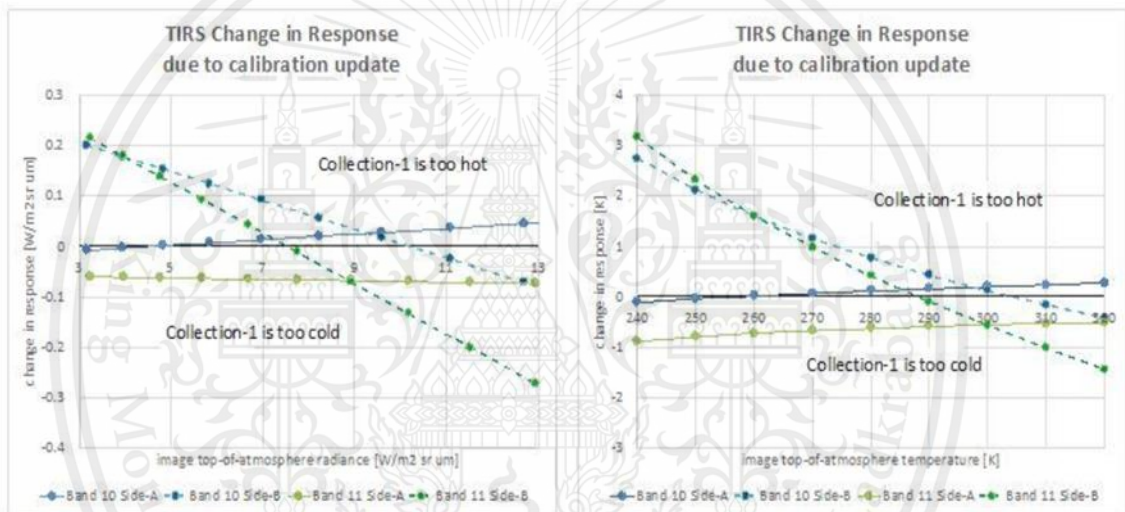


Figure 3.3 The images display the estimated change of Landsat 8 Thermal band calibration update

(Source:<https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8-oli-and-tirs-calibration-notice>)

- Landsat 8 TIRS Relative Gains:** To adjust for striping in Landsat Collection 1 picture data, the detector-to-detector radiometric calibration of both Landsat 8 TIRS bands was revised. Radiometric striping has increasingly deteriorated over the mission's existence, according to data from the on-board calibrator. The **Figure 3.4** shown to the right compare the striping that is evident in TIRS Band 10 with Collection 1, with noticeable reduction in the Collection 2 image. (Data acquired December 26, 2019.)

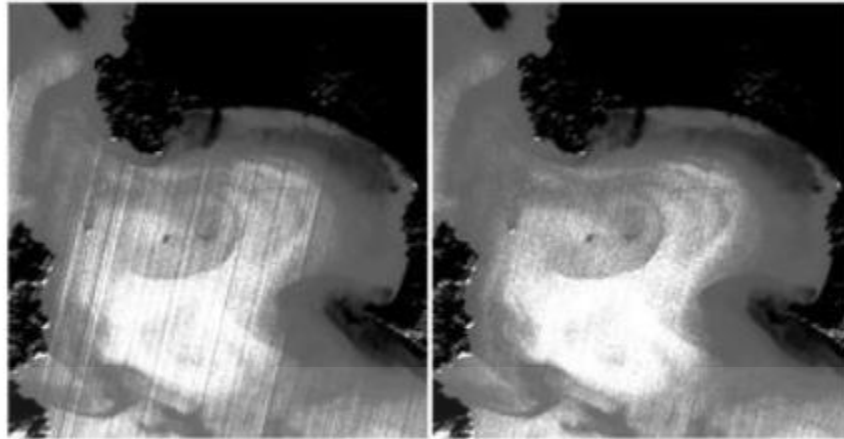


Figure 3.4 Left: Sample of a TIRS Band 10 image with radiometric striping when processed into Collection 1 and Right: The reduction of striping in Collection 2

(Source: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-1-data?qt-science_support_page_related_con=1#qt-science_support_page_related_con)

- **Landsat 8 Absolute Gain Model Change:** The radiometric calibration model for the Landsat 8 OLI is based on an average response of the on-board calibrators (three lamp pairs, two solar diffuser panels and lunar acquisitions). The data from the working lamp pair, used daily, and from the working solar diffuser panel, used weekly, show trends that disagree with the other calibrators, indicating that their characteristics are changing independently of the instrument. The calibration model was updated to no longer include the working solar diffuser and the working lamp pair. Only the Coastal Aerosol (Band 1) and Blue (Band 2) are affected by this change.
- **Landsat 8 OLI Relative Gain Update:** The variations between Collection 1 and Collection 2 photos are typically less than 0.2 percent; however, some individual detectors in the SWIR bands can alter by up to 2.0 percent. The relative improvements will continue to be adjusted in future quarters based on predictions from solar diffuser data from the previous quarter.
- **Landsat 8 OLI Bias Calculation Update:** The method of calculating the Landsat 8 OLI bias, used for dark subtraction in the calibration process, was modified to make use of Video Reference Pixels (VRPs). The improved bias estimates reduce along-track noise, improve image uniformity, and improve signal-to-noise ratio. The effect of this update is primarily noticeable in low radiance scenes, as shown in **Figure 3.5** example to the right. The reduction in striping is a result of the improvement.

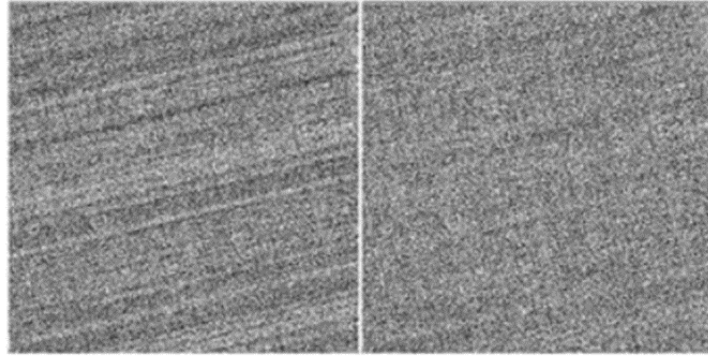


Figure 3.5 Landsat 8 OLI images of collection 1 on the left and collection 2 on the right

(Source:https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-1-data?qtscience_support_page_related_con=1#qt-science_support_page_related_con)

3.2.2 Landsat collection 2 Level-2

Information

Landsat Level-2 science products are generated from Collection 2 Level-1 inputs that meet the <76 degrees Solar Zenith Angle constraint and include the required auxiliary data inputs to generate a scientifically viable product. The **Figure 3.6** below will show about different between Landsat collection 2 examples.

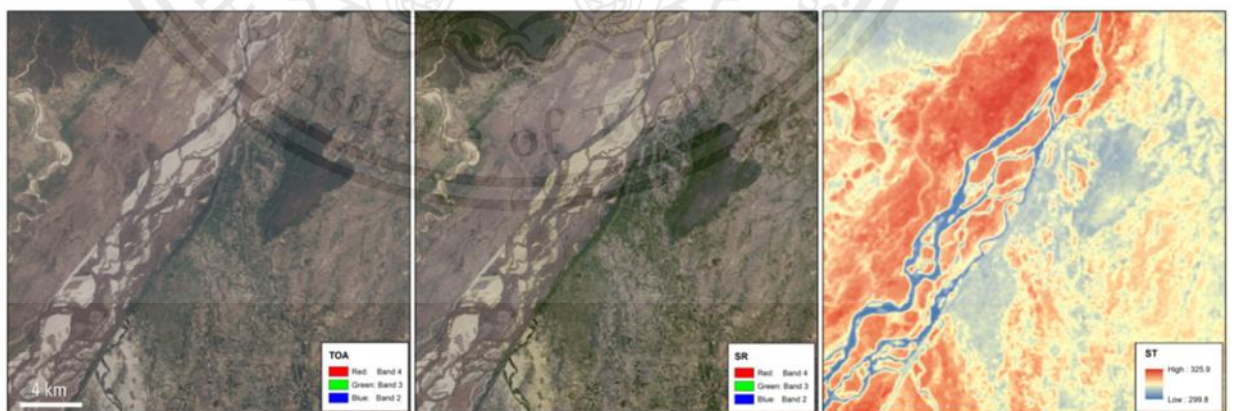


Figure 3.6 Landsat collection 2 examples

(Source:<https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products>)

From the left, the image is Landsat 8 Collection 2 level-1 top of atmosphere reflectance image, Landsat 8 Collection 2 level-1 top of atmosphere reflectance image and Landsat 8 Collection 2 level-2 surface temperature image for an area over the Sapta Koshi River in Bairawa, Nepal (path 140 row 41) acquired on May 3, 2013.

Surface Reflectance

The percent of incoming solar radiation reflected from the Earth's surface to the Landsat sensor is measured by surface reflectance (unitless). The LEDAPS and LaSRC surface reflectance algorithms account for the atmospheric gases, aerosols, and water vapor's temporally, geographically, and spectrally variable scattering and absorbing effects, which is required to accurately define the Earth's land surface. Some specific enhancements are shown in **Table 3.2** below.

Table 3.2 Collection 2 Surface Reflectance Enhancements

Collection 2 Surface Reflectance Enhancements		
	Collection 2 Surface Reflectance	Collection 1 Surface Reflectance (on demand, U.S. ARD)
Fill Value	0	-9999
Scaling Factor	0.0000275 + -0.2	0.0001
Data Type	Unsigned 16 bit integer	Signed 16 bit integer
Valid Range	1-65455	0-10000
Quality Band	L1 QA_Pixel L1 QA_RADSAT L2 SR_QA_AEROSOL	Level-2 PIXELQA Level-2 RADSATQA Level-2 SRAEROSOLQA

(Source:<https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products>)

Surface Temperature

Surface Temperature measures the Earth's surface temperature (units Kelvin) and is an important geophysical parameter in global energy balance studies and hydrologic modeling. Surface temperature is also important for monitoring agricultural

and vegetation health, as well as extreme heat events like volcanic eruptions and wildfires, as well as urban heat island effects. A surface temperature product cannot be generated if auxiliary data is missing, although a surface reflectance product will continue to be generated. Some specific enhancements are shown in **Table 3.3** below.

Table 3.3 Collection 2 Surface Temperature Enhancements

	Collection 2 Surface Temperature	Collection 1 Surface Temperature (on demand, U.S. ARD)
Fill Value	0	-9999
Scaling Factor	0.00341802 + 149.0	0.1
Data Type	Unsigned 16-bit integer	Signed 16-bit integer
Valid Range	1-65535	1500-3730
Atmospheric Reanalysis Source	GEOS-5 FP-IT (2000-present) MERRA-2 (1982-1999)	NARR
Quality Band	L1 QA_Pixel L1 QA_RADSAT L2 SR_QA_AEROSOL	Level-2 PIXELQA Level-2 RADSATQA Level-2 SRAEROSOLQA

(Source: <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products>)

Atmospheric Auxiliary Data

Landsat Level-2 surface reflectance and surface temperature processing require atmospheric auxiliary data input from external USGS data sources. USGS retrieves data from the data source and extracts parameters specific to Landsat Collection 2 Level-2 processing. These subsets are available for download for users who wish to perform Level-2 processing. It is not necessary for users to download atmospheric auxiliary data to use the Collection 2 Level-2 products.

3.3 Comparison Landsat 8 collection 1 and 2

There are many updating from Landsat collection 1 to collection 2 that benefit to our project so lastly, we chose Landsat collection 2 instead of collection 1. From the table 3.4, shown a comparison between Landsat 8 collection 1 and collection 2.

Table 3.4 Comparison between Landsat 8 collection 1 and 2

	Collection 1		Collection 2		
	Level-1	Level-2 U.S.	Level-1	Level-2	Level-2 U.S.
Supporting software					
Processing software	Landsat Product Generation System (LPGS) R13.x	L2PGS BRIDGE R0.8.x	Landsat Product Generation System (LPGS) R15.x		
Geometry					
Coverage	Global	U.S	Global	Global	U.S
Geometric Registration Base	Global Land Survey (GLS) 2000		Landsat 8 OLI Harmonized w/ Sentinel-2 Global Reference Image (GRI)		
Digital Elevation Model	GLS DEM (SRTM / NED / CDED / DTED / NPI / GIMP / RAMP)**		GLSDEM / NASADEM / Alaska NED / CDEM / SNF / NPI / GIMP / ArcticDEM / RAMP**		
Precision Correction Methodology	Baseline		Improved usage of Ground Control Points (GCPs) to produce more Level-1 Terrain Precision (L1TP) Products		
Radiometry					
TIRS Post-Stray Light Correction Adjustment (L8 Only)	None		Post-Stray Light Residual Bias Applied	Post-Stray Light Residual Bias Applied	
Atmospheric Correction / Level-2 Processing					
Surface Reflectance Algorithm Version	N/A	LEDAPS v3.2.1 (TM / ETM+) LaSRC v1.3.0 (OLI / TIRS)	N/A	LEDAPS v3.4.0 (TM / ETM+) LaSRC v1.5.0 (OLI / TIRS)	
Surface Reflectance Fill Value	N/A	-9999	N/A	0	
Surface Temperature Fill Value	N/A	-9999	N/A	0	
Data Type / Scaling Factor (Surface Reflectance)	N/A	Signed 16-bit integer 0.0001 (no offset)	N/A	Unsigned 16-bit integer 0.0000275 + -0.2	
Data Type / Scaling Factor (Surface Temperature)	N/A		N/A	Unsigned 16-bit integer 0.00341802 + 149.0	
Atmospheric Reanalysis Source (Surface Temperature)	N/A	North American Regional Reanalysis (NARR)	N/A	GEOS-5 FP-IT (new data) / MERRA-2 (archive data)	
Pixel-Level Metadata					
Cloud Masking	Cloud, Cloud Confidence, Cloud Shadow Confidence, Cirrus Confidence (L8 only)		Cloud, Cloud Confidence, Cloud Shadow, Cloud Shadow Confidence, Dilated Cloud, Cirrus (L8 /L9 only), Cirrus Confidence (L8 /L9 only)		
Land Characterization Bits	Snow / Ice Confidence		Snow, Snow / Ice Confidence, Water		
Radiometric Saturation	High-Level (Number of Impacted Bands Only)		Per-Band Saturation Bits		
Product Format / Packagin					
Format	GeoTIFF		Cloud-Optimized GeoTIFF (COG)		
Band Subsetting	All Bands Included in Product		User-Selectable Bands		
Data Access / Visualization Tools	EarthExplorer Glovis LandsatLook Viewer Machine-to-Machine (M2M) Bulk Download		EarthExplorer Machine-to-Machine (M2M) Bulk Download Direct Access via AWS S3 ^^		

(Source: <https://www.usgs.gov/media/files/landsat-collection-1-vs-collection-2-summary>)

From the topic 3.2 and 3.3, the data shows that Landsat collection 2 have a huge improvement far from Landsat collection 1 whether it is accuracy, ground control point, error improvement and new data such as atmospheric auxiliary data. The most important function for our project is surface reflectance which mean the raw data will not include atmospherics. To conclude Landsat collection 2 are more accurate and suitable for our project since the project aims to find a vegetations area on earth surface.

3.4 Confusion Matrix

A confusion matrix is a table that often used to describe the performance of a classification model on a set of test data which the true values are known. (In our case the true value called ground truth).

Table 3.5 Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positive (TPs)	False Positive (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

The most basic terms of the matrix are true positives, true negatives, false positives, false negatives, all these terms are whole numbers.

- True positives (TP): These are cases in which we predicted yes, and the answer is yes.
- True negatives (TN): We predicted no, and the answer is no.
- False positives (FP): We predicted yes, but the answer is no.
- False negatives (FN): We predicted no, but the answer is yes.

With these data from the confusion matrix, we can compute accuracy, misclassification, precision, recall, F-score.

Accuracy: Accuracy is the number of correctly predicted results out of the total accuracy should be considered when TP and TN are more important than the others.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3.4)$$

Precision: Out of the total predicted positive values, how many were positive.

$$\text{Precision} = TP / (TP + FP) \quad (3.5)$$

Recall: Out of the total actual positive values, how many were correctly predicted as positive.

$$\text{Recall} = TP / (TP + FN) \quad (3.6)$$

F-1 score: When FP and FN both are equally important. This allows the model to consider both precision and recall equally using a single score.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.7)$$

Conclusion

The confusion matrix is a useful tool for our project since the matrix is the main information's format using to measure the performance of a classification models. We can define our classification models precision, recall and effectively accuracy. With this information we can improve the model if the accuracy is not satisfied.



CHAPTER 4
RESEARCH METHODOLOGY

Overview of Research Methodology

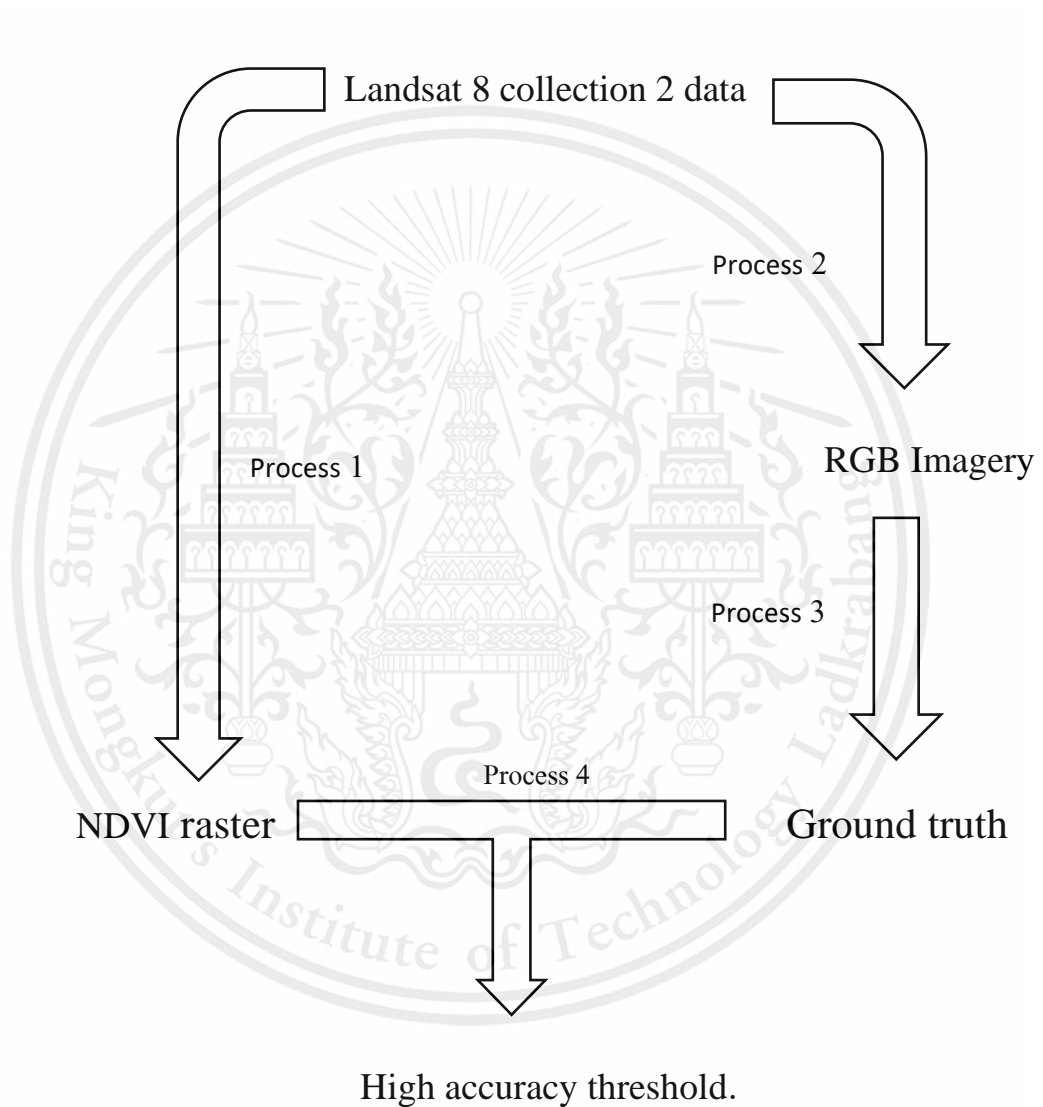


Figure 4.1 Research methodology overview

Figure 4.1 shown input, output, and overview of research methodology. As the chart shown satellite imagery was used at the initial phase of our research to provide NDVI raster and RGB image through process 1 and 2. After we obtain RGB image, process 3 are in charge to produce ground truth. In last process ground truth and NDVI raster are required for processing to obtained high accuracy threshold.

- Process 1: Finding NDVI.
Since our project aims to find a green area. In this process we convert raw data from Landsat8 collection 2 into an NDVI raster. Firstly, we have raster data from USGS in tiff format. The output of this process is a NDVI imagery in tiff format. The NDVI value will vary around -1 to 1 in every pixel.
- Process 2: Converting to RGB.
In this process we convert raw data from Landsat8 collection 2 into RGB imagery which combined red, green, blue colors to produce a broad array of colors in tiff formats.
- Process 3: Digitizing.
Digitizing process we manually digitize information (green area in our case) that we want into tiff format. We expected ground truth form this process.
- Process 4: Finding proper threshold.
This process we are looking for threshold that satisfied the accuracy rate. This project is crucial since the project are about the accuracy of the prediction. The process involves the experiment called 3 type of area which we compare the maximum accuracy threshold of a different types of area. The study area is rural, urban and forest area.

4.1 Finding NDVI.

This process we convert raw data into NDVI image using NDVI formula 3.3 since we used Landsat 8 as a source data and USGS separate their raw data into a band. So, we can calculate NDVI directly via this formula. The result of this process is a raster data that contain NDVI value in each pixel. The value will vary in -1 to 1. If value in the pixels is close to 1 it tends to be a vegetation area and other is not. In **Figure 4.2** shown NDVI image that composite form NIR and Red bands.

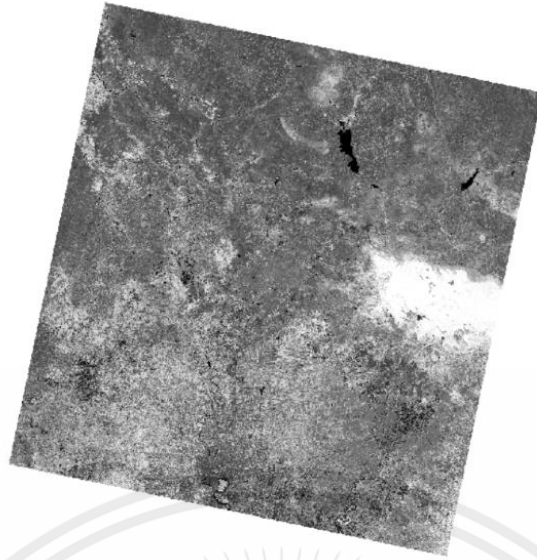


Figure 4.2 NDVI image converting form LANDSAT 8 collection2

4.2 Converting to RGB.

Color digital image are made of pixels, and each pixel made of combination of primary colors represented by series of code. Bands or Channels are containing different information include primary colors which are red, green, blue. The pixel in each band is contain number value. A band in this case is extended to be the greyscale imaged based on any such conventional primary color. An RGB image as you can see in **Figure 4.3** has three bands Red, Green, Blue. RGB channels roughly follow the color receptors in the human eye and are used in computer displays. In other word RGB image is used to present an image in regular way. In this process we are stacking red, green, blue band together to get a RGB image. The importance of these process is to use RGB in digitizing process.



Figure 4.3 RGB image

4.3 Digitizing

Digitizing process is to classify the green area manually. The result is the representation of a ground truth (**Figure 4.5**). We used the results of this process to compare with vegetation prediction from our algorithm. Firstly, we clipped RGB image into smaller pieces as you can see in **Figure 4.4** and create polygon to cover the green area. After that we got a result of green area in area polygon format. In digitizing process, we create polygon pixel by pixel to minimize an error. (**Figure 4.5**) To compare the ground truth to our prediction, we need to convert into raster tiff format because polygon format is just an area. But raster formats are containing indicated value. In our project we use value 1 to indicate a green area and 0 if it is not.



Figure 4.4 Clipped RGB imagery



Figure 4.5 Digitizing process

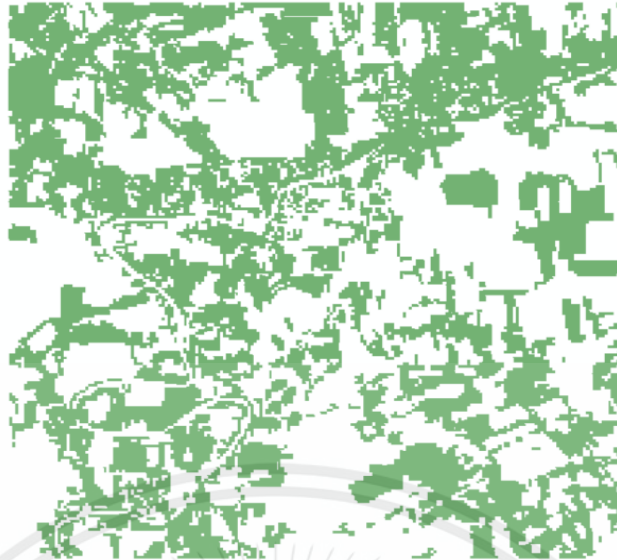


Figure 4.6 Raw ground truth form digitizing process.

A raw ground truth as you can see in **Figure 4.6** is in a polygon format. Therefore, converting into a raster format are required. Converting are straight forward, using a tool in ArcGIS called polygon to raster and ground truth are ready.

4.3.1 Compare Ground truth

After we got ground truth from digitizing process, we must check if the raster is trustable. By comparing ground truth from two people that digitize the same area and compare the results. The parameter of these experiment is percent's overlap and percent's accuracy. The process required a ground truth of the same area that obtain by two or more people. In our study we use two candidates to obtain a ground truth. **Figure 4.7** and **Figure 4.8** are the ground truth that provide by two people.

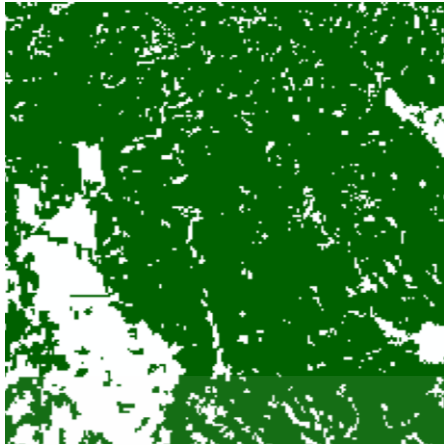


Figure 4.7 Ground truth 1 (Koltouch)
(Marwin)



Figure 4.8 Ground truth 2

After we obtained two specimens from two different people, we can create an overlap area. An overlap area in **Figure 4.9** used to guide the authenticity of the ground truth.



Figure 4.9 Area overlap ground truth.

4.4 Finding proper thresholds.

In this process we aim to find a proper threshold that can satisfied a maximum accuracy of the prediction. To find a proper threshold, ground-truth and NDVI and the most importance confusion matrix. Both ground truth and NDVI are required in raster formats. Firstly, we create a confusion matrix to describe the quality of the prediction by provide us an accuracy.

4.4.1 Confusion Matrix

Table 4.1 Confusion Matrix

	Ground truth (1)	Ground truth (0)
Predicted (1)	True Positives (TP)	False Positive (FP)
Predicted (0)	False Negative (FN)	False Positive (FP)

Normally confusion matrix is required two information input called true class and prediction class. True class are the class that we used as a reference of the prediction. Predicted class are the class we predicted by our classification algorithm. In our project we apply ground truth as a true class and NDVI classification as a prediction class as **Table 4.1**. The number zero and one after Ground Truth and Predicted shown the output of those prediction and NDVI, 1 means It is this area are green area and 0 means these areas are not. The example of classified NDVI raster is shown in **Figure 4.10**.

1	1	1	1	1	1	1	1	1	1	1	0
1	0	0	0	1	0	1	1	1	1	1	1
1	1	1	0	0	1	1	1	1	1	1	1
1	1	1	1	0	1	1	1	1	0	1	1
1	1	1	1	1	0	0	0	1	0	0	1
1	1	1	1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	0	0	0	0	1	1
1	0	1	1	1	1	0	0	0	0	1	1
1	0	0	1	1	1	1	1	0	1	1	1
1	1	1	1	1	0	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1

Figure 4.10 Example of classified NDVI raster

Since we only have one ground truth for each image since it is a reference. In the other hand NDVI prediction can will vary. After we got NDVI image form process 1 we create 100 arrays that contained threshold value. The value must in the range of -1 and 1. So, each array will contain the same value in each pixel. We called it threshold classifier array.

0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

Figure 4.11 Examples of NDVI threshold classifier

Figure 4.11 shown some examples of classifier array. As you can see all the pixels value are the same in every array. These arrays will have x, y dimensions as same as the NDVI raster. We provide a condition form these classifier arrays to create a NDVI classified array. The NDVI classified array is an array that contain value number only -1 and 1. Since the NDVI array contains floats value (the number that have decimals place). The condition we used to create NDVI classified is.

If NDVI arrays \geq classifier arrays; NDVI classified = 1

If NDVI arrays $<$ classifier arrays; NDVI classified = -1

After we get a classified array for all threshold values. We are looking for a threshold that satisfied the highest accuracy.

4.4.2 Finding Accuracy.

To find an accuracy of a classified array for each threshold. The confusion matrix in the beginning of this chapter is required. Accuracy in our condition means how well the prediction is in other words how well the prediction is out of all prediction. We can apply this formular to the confusion matrix since confusion matrix give us a True Positives, True Negatives, False Positives, False Negatives.

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{Total predictions}} \quad (4.1)$$

Where Correct predictions = true positives + true negatives

Total prediction = true positives + true negatives + false positives
+ false negatives

With this equation we can find an accuracy for each threshold and plot and looking for the one that satisfied the maximum. Form the formula, to find accuracy TP, TN, FP, FN are required, so confusion matrix is in charge. After we get TP, TN, FP, FN form the matrix. Then we can calculate the NDVI and plot.

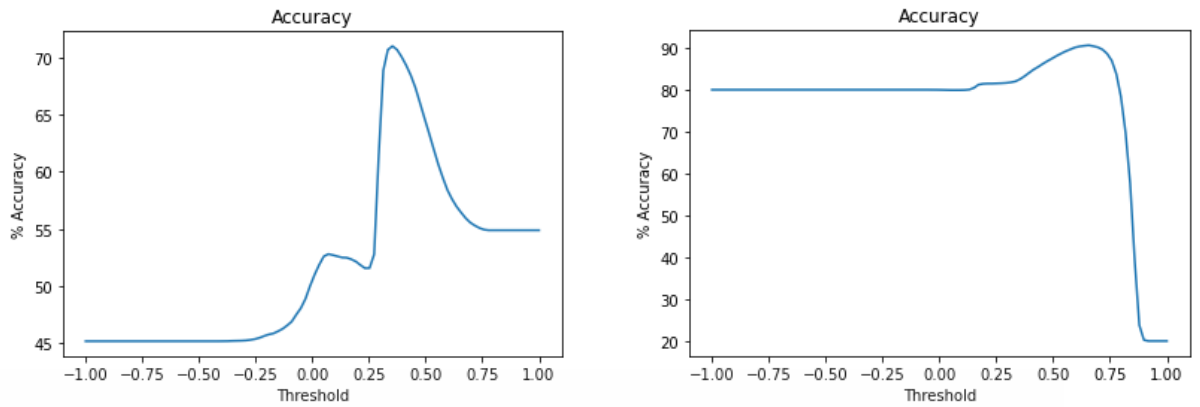


Figure 4.12 plotting (Threshold, Accuracy%)

Figure 4.12 is the result of an accuracy and NDVI. Both graphs come from the same satellite image. In the first graph threshold that satisfied maximum accuracy is around 0.375 and the maximum accuracy is around 70 percent. In the second graph threshold that satisfied maximum accuracy is around 0.7 and the maximum accuracy is 90 percent. Although both graphs are come from the same image, but the outcome are not having to be the same. Even though we can find suitable threshold for each graph, the results do not satisfy. Because we need threshold that provide high accuracy in every area. So, we studied furthermore to find the best possible threshold using these studying parameters.

- Area Types (Urban, Rural, Forest)
- Temporal Area (Varying date and time that image was taken)

Both were used as a parameter to observe of a threshold that satisfied high accuracy and studies for an effect in thresholds form different situations.

4.4.2.1 Area Types

The area type process studies are to compare the result of each image that come with the different types of area. The studies include forest area, rural area, and urban area. We expected to see if the different of area are affect the proper threshold and looking for threshold that satisfied high accuracy for all images. The process started with raw imagery form Landsat 8 collection2. After that we plot an accuracy graph for each type of area and compare the results. All specimens are clipped form the same satellite image. The properties of clipped images are the same with 30meters cell size and 200*200 pixels.

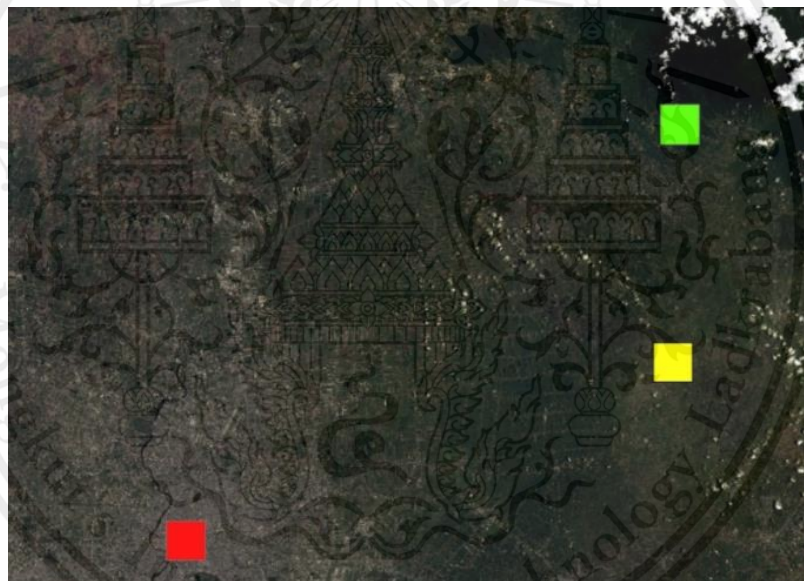


Figure 4.13 Overview of area, each square represented different types of terrain

Figure 4.13 shown overview of an area that we clipped form satellite imagery. The red square represented an urban area as you can see in **Figure 4.14**. The yellow square

represented the rural area in **Figure 4.15**, and the green square represented the forest area as shown in **Figure 4.16**.



Figure 4.14 Urban area

Urban area is the region surrounding a city. Urban areas are very developed, meaning there is a density of human structures such as houses, commercial buildings, roads, bridges, and railways.



Figure 4.15 Rural area

Rural areas are areas which are not towns or cities. They are often farming or agricultural areas.



Figure 4.16 Forest area.

Forest area is land under natural or planted stands of trees. After we got all this area that represented different terrains, we apply the previous process to provide a proper threshold for each clipped area. We decided to apply the process to two satellite images, then we compare the results. **Figure 4.17, Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21, and Figure 4.22** show the graph that plot form threshold and accuracy of both scenes, to observe for thresholds that maximize an accuracy.

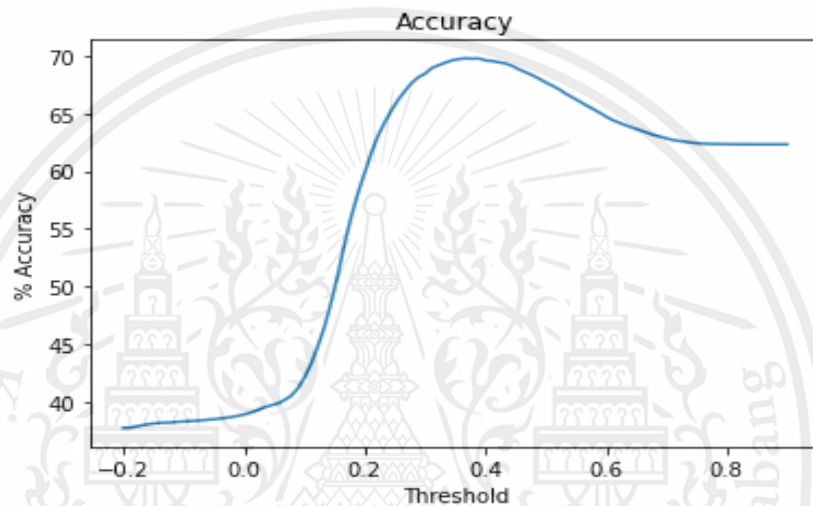


Figure 4.17 Plot of accuracy and thresholds of Urban area scene 1.

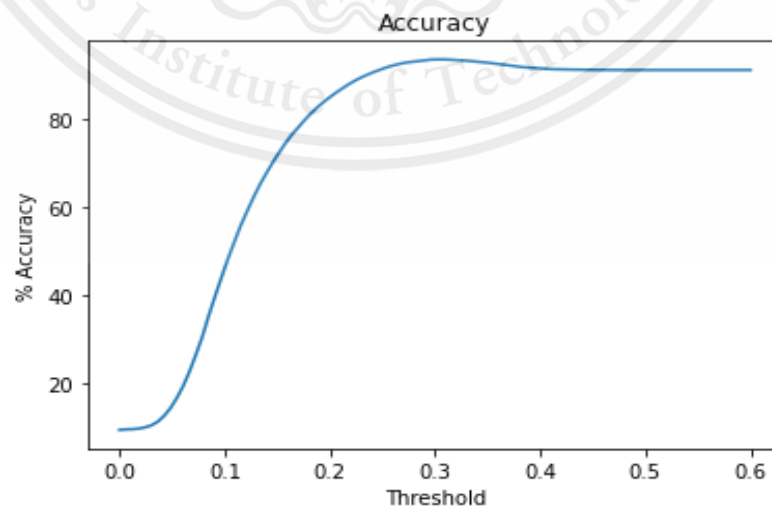


Figure 4.18 Plot of accuracy and thresholds of Urban area scene 2.

This material is reserved for educational use only, not allowed for commercial use.

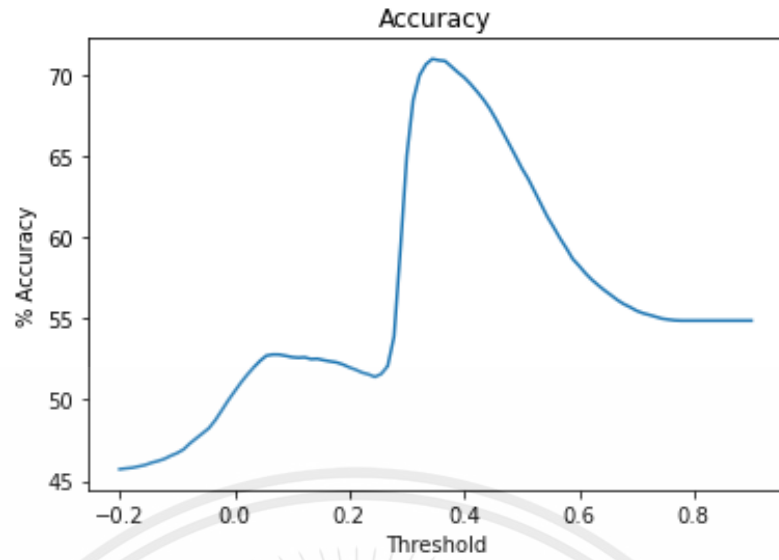


Figure 4.19 Plot of accuracy and thresholds of Rural area scene 1

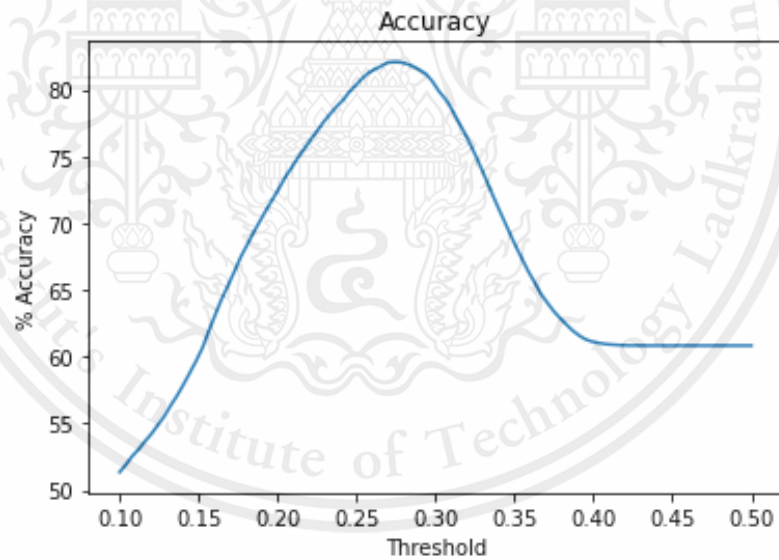


Figure 4.20 Plot of accuracy and thresholds of Rural area scene 2

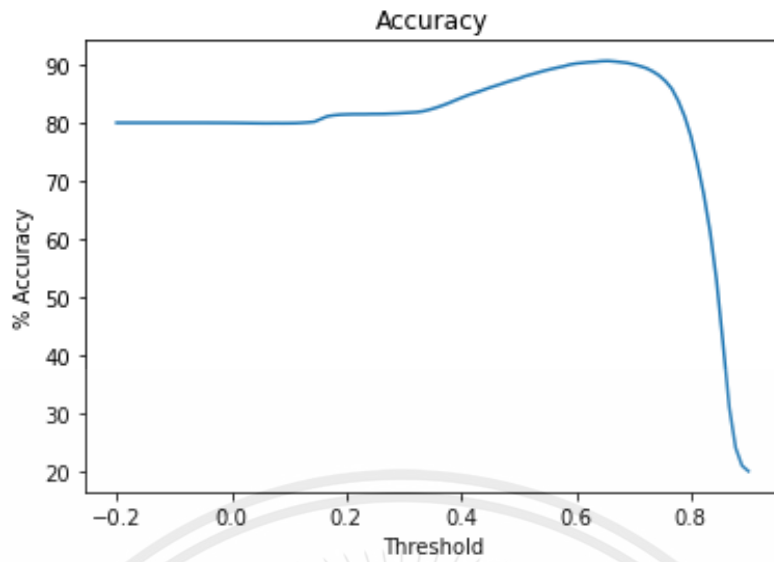


Figure 4.21 Plot of accuracy and thresholds of Forest area scene 1

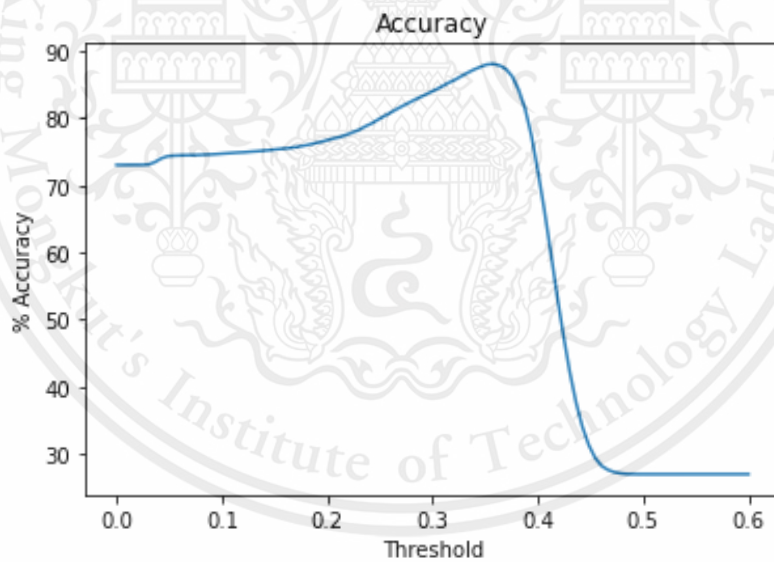


Figure 4.22 Plot of accuracy and thresholds of Forest area scene 2

Figure 4.23, Figure 4.24, Figure 4.25 shown the prediction of each area type in scene 1. To maximize accuracy in scene 1 threshold are 0.367,0.372, 0.655 for urban, rural and forest in order. **Figure 4.26, Figure 4.27, Figure 4.28** shown the prediction of each area type in scene2. To maximize accuracy in scene 2 threshold are 0.303, 0.274, 0.358 for urban, rural and forest in order.

Maximum Acc predicted NDVI(0.367 Threshold)

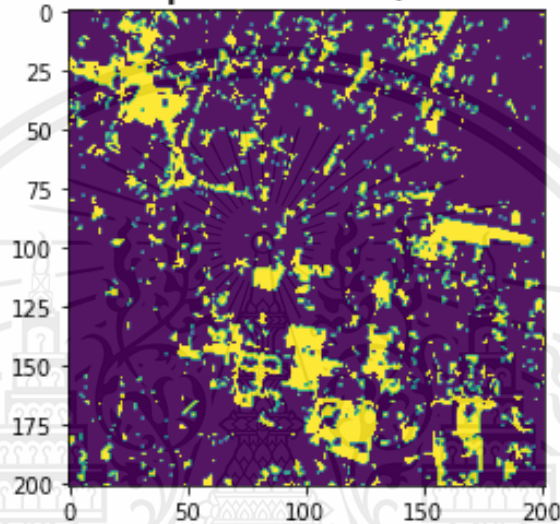


Figure 4.23 Green area Prediction with maximum accuracy of urban area scene 1

Maximum Acc predicted NDVI(0.372 Threshold)

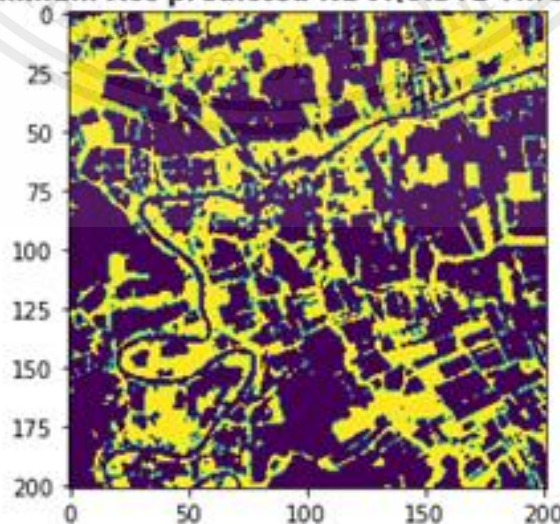


Figure 4.24 Green area Prediction with maximum accuracy of rural area scene

This material is reserved for educational use only, not allowed for commercial use.

Maximum Acc predicted NDVI(0.655 Threshold)

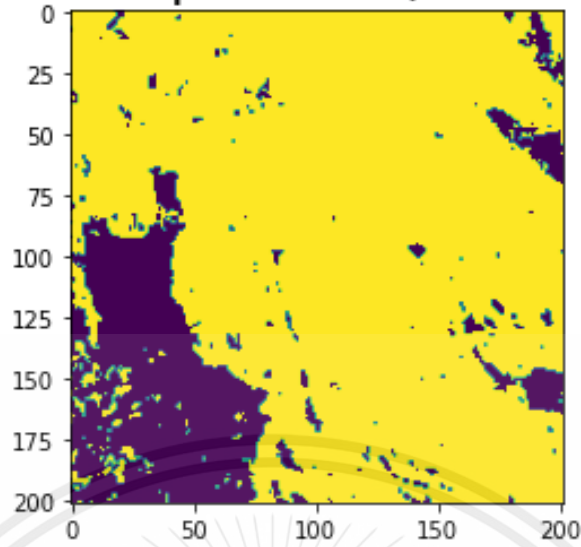


Figure 4.25 Green area Prediction with maximum accuracy of forest area scene 1

Maximum Acc predicted NDVI(0.30303 Threshold)

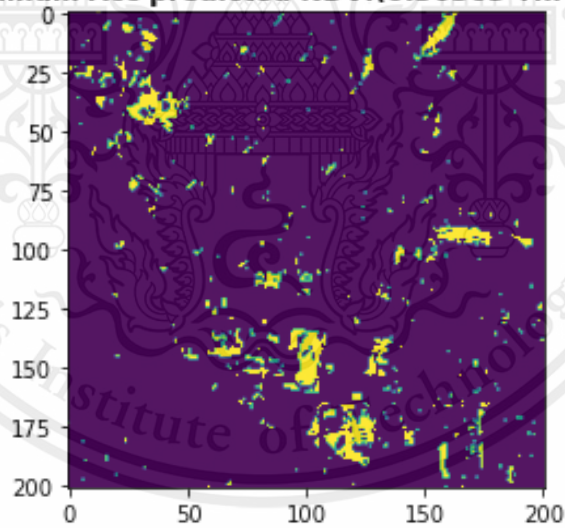


Figure 4.26 Green area Prediction with maximum accuracy of urban area scene 2

Maximum Acc predicted NDVI(0.2737 Threshold)

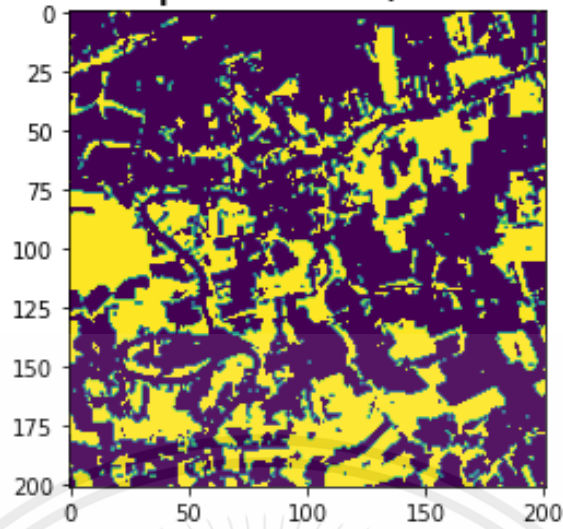


Figure 4.27 Green area Prediction with maximum accuracy of rural area scene 2

Maximum Acc predicted NDVI(0.357576 Threshold)

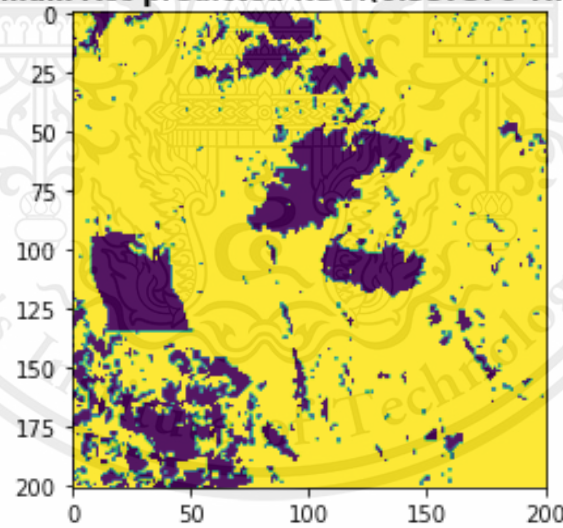


Figure 4.28 Green area Prediction with maximum accuracy of forest area scene

4.4.2.2 Temporal Area

The temporal area we study an area with different timing. In other word we are finding a proper threshold of the same area with different time to observe if there are a differ in threshold and compare the result to find a threshold that work well. This process aims to observe for a threshold that suit in a particular area. The process will be started with a particular area. Then we download raw satellite image of the same area with different timing. After that we find a proper accuracy for each image and compare the results. The process started with image of the same area which taken in a different timing as you can see in **Figure 4.29, Figure 4.30, Figure 4.31.**

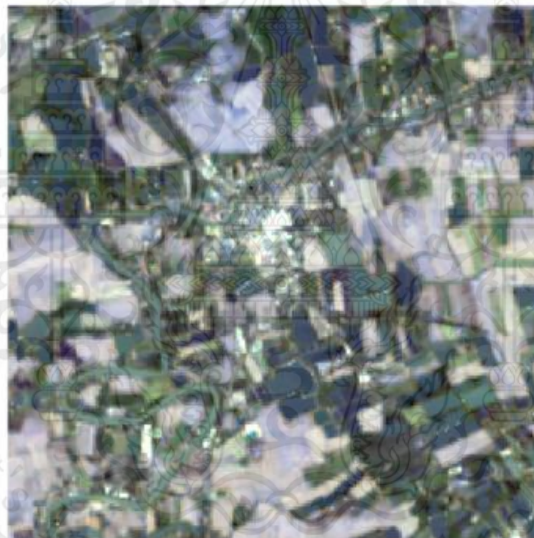


Figure 4.29 RGB image converted form satellite image taken in 19/2/2020



Figure 4.30 RGB image converted form satellite image taken in 29/8/2020



Figure 4.31 RGB image converted form satellite image taken in 17/11/2020

After we got all these images, we started to find the proper for each image. The maximum threshold is shown in **Table 4.2**.

Table 4.2 Table of maximum accuracy and thresholds form temporal experiment

Results	Date		
	19/02/2020	17/11/2020	29/08/2020
Maximum accuracy	71.03%	81.386	74.06
Threshold	0.372	0.342	0.261



CHAPTER 5

RESULTS AND DISCUSSTIONS

5.1 Results

From the experiment that we used area types and temporal area experiment to observe for a threshold that satisfied high accuracy we found that although NDVI can suggested a green area by using reflected light in the visible and near-infrared bands. To obtained trustable results ground truths area relevant since we used it as a reference for both experiments. The results provide not only a threshold form both 3 types of area experiment and temporal area experiment but also the results of overlap area form comparing ground truth process.

5.1.1 Three types of area result.

As you can see all three images in the first scene has a different threshold that satisfied the maximum value of accuracy. For urban area to reach a maximum accuracy threshold equal to 0.367, the maximum accuracy is 70.12%. For rural area, threshold equal to 0.372 to reach maximum accuracy, the maximum accuracy is 71.03%. Lastly the forest area, to satisfy maximum accuracy of 90.65% it is required threshold equal to 0.655. For the second scene, the maximum accuracy of urban area, rural area and forest area are 93.61%, 82.08%, 88.09% and threshold that satisfied maximum accuracy is 0.303 for urban area, 0.274 for rural area and 0.357 for forest area. Form the results form area type experiment as shown in **Table 5.1** that we have done. Means accuracy that satisfied the maximum accuracy is 0.4647 for first scene and 0.307 for second scene. To find a proper threshold that can satisfied all area type, using mean values does

not work out. Since 0.4647 and 0.307 threshold cannot maintain the high percent of accuracy. So, we decided to use a threshold form the area that give us lowest accuracy compared to other. That is a threshold form urban area scene 1. At 0.367 threshold satisfied us in term of overall accuracy which equal to 77.726%. Form the results we apply to a satellite image to represented green area prediction at threshold equal to 0.367 as shown in **Figure 5.1**.

Table 5.1 Results of Area types experiment

Results	Area Types					
	Urban (1)	Rural (1)	Forest (1)	Urban (2)	Rural (2)	Forest (2)
Maximum accuracy	70.12%	71.03%	90.65%	93.61%	74.06%	88.09%
Threshold	0.367	0.372	0.655	0.303	0.261	0.357
Threshold (0.367)	70.12%	70.90%	83.97%	92.36%	64.71%	84.63%



Figure 5.1 Green area prediction at 0.367 threshold compared to RGB

5.1.2 Temporal area experiment.

From the experiment results at threshold equal to 0.372,0.342,0.261 with an accuracy of 0.372, 81.386, 74.06 for an image that was taken on 19/02/2020, 17/11/2020, 29/08/2020. To observe a threshold that work well for all timing we use each threshold and mean of all three-timing scene to find one proper threshold that satisfied overall high accuracy.

Table 5.2 Results of temporal area experiment

Threshold		Accuracy %			
		19/02/2020	17/11/2020	29/08/2020	Mean accuracy
19/02/2020	0.372	70.723	81.1044	64.0454	71.9576
17/11/2020	0.342	70.996	81.3866	70.4488	74.27713333
29/08/2020	0.261	51.766	75.81	81.5326	69.70286667
Mean threshold	0.325	70.4364	80.93	74.8348	75.4004

From **Table 5.2** means accuracy of the images is 71.958, 74.277, 69.703 and at mean of all three image, threshold equal to 0.325 and provide accuracy of 75.4 which is the highest compared to other. **Figure 5.2, Figure 5.3, Figure 5.4** shown the results prediction of green area that vary by timing using 0.325 threshold.

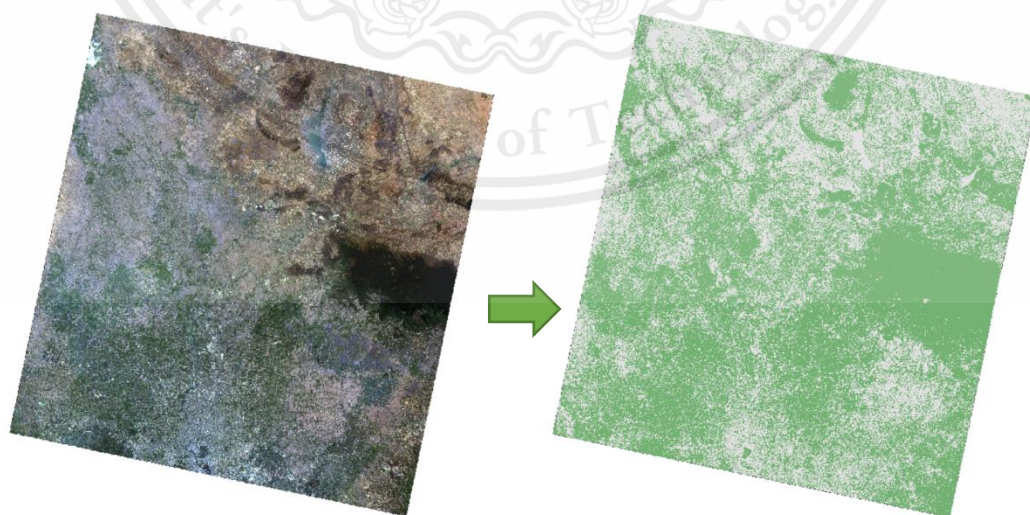


Figure 5.2 Green area prediction at 0.325 threshold compared to RGB (Taken at 19/02/2020).

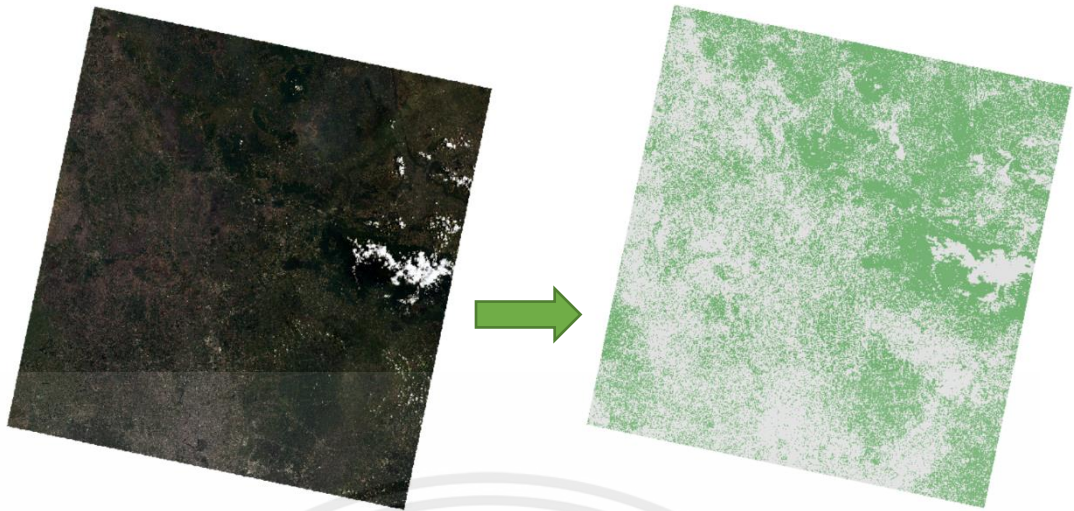


Figure 5.3 Green area prediction at 0.325 threshold compared to RGB (Taken at 29/08/2020)



Figure 5.4 Green area prediction at 0.325 threshold compared to RGB (Taken at 17/11/2020)

5.1.3 Compare ground truth results.

From digitizing ground truth 1 cover 84.381% of all areas and 80.022 for ground truth 2. Ground truth 1 and 2 reach maximum accuracy of 90.648% and 95.490% at thresholds of 0.655 and 0.633. **Figure 5.4** shown the area of both ground truths 1 and 2 that overlap form each other. An overlap area contains 97% of ground truth 1 and 92% of ground truth 2.

5.2 Discussions

The observing of this study suggests that using NDVI threshold are effective enough to classified for satellite image. However, our studied has shown more information's that cause differ in threshold's accuracy. Since threshold of both thresholds form area types and temporal area experiments are differ from each other's, it shown that types of terrain and timing that images was taken are affected the results of threshold's accuracy.

CHAPTER 6

CONCLUSIONS & FUTURE WORK

6.1 Conclusions

Green area is one of the most important area to concerns since climate change and deforestation issues have a big impact on humans and environments. To measure the green area are directly relevant to the problems.

In this study, satellite image form Landsat 8 collection 2 was used as a source of data. different area types and temporal experiments were used to observe for a threshold that satisfied maximum accuracy. Form both experiments. The effective threshold form area type experiment is 0.367 and 0.325 for temporal experiment. Both can satisfy high value of accuracy which is 77.782% for area type and 75.4% for temporal area. As the results we found that even we got a NDVI image of the same area the prediction of a green area will change since vegetation are not permanent.

6.2 Future work

Many experiments, tests, samples have been left behind for the future due to the lack of time. Some processes are taking a lot of time such as digitizing process, to provide a precisely ground truth it required a lot of time and effort. In future we aim to go deeper to observe for a threshold that can satisfied 90% of accuracy. To develop further in term of accuracy more area and ground truth quality are relevant. Applying our work to find green area are significantly benefit for many associations such as Forest department, Ministry of Agriculture. More than that these green area observations can increase awareness of deforestation and climate change issues.

REFERENCES

- Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, 34, 136-146.
- Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very high-resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE geoscience and remote sensing letters*, 15(4), 607-611.
- Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., ... & Moghaddam, M. (2014). A python-based open-source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables. *Remote Sensing*, 6(7), 6111-6135.
- Taufik, A., Ahmad, S. S. S., & Ahmad, A. (2016). Classification of Landsat 8 satellite data using NDVI thresholds. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(4), 37-40.

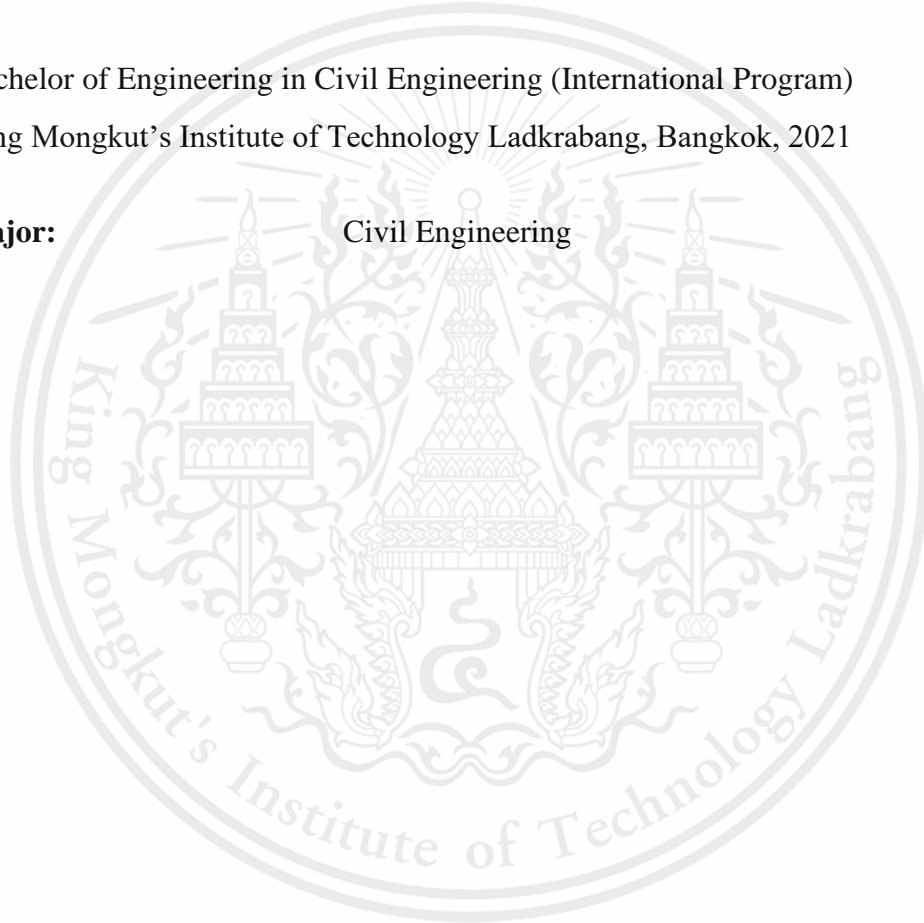
AUTHOR BIOGRAPHY

Author: Mr. Koltouch Anantakarn
Degree: Bachelor of Engineering
Date: 4th June 2021
Date of Birth: 21st September 1998
Place of Birth: Bangkok, Thailand

Undergraduate Education:

Bachelor of Engineering in Civil Engineering (International Program)
King Mongkut's Institute of Technology Ladkrabang, Bangkok, 2021

Major: Civil Engineering



AUTHOR BIOGRAPHY

Author: Mr. Marwin Suwanprapa
Degree: Bachelor of Engineering
Date: 4th June 2021
Date of Birth: 31st October 1998
Place of Birth: Bangkok, Thailand

Undergraduate Education:

Bachelor of Engineering in Civil Engineering (International Program)
King Mongkut's Institute of Technology Ladkrabang, Bangkok, 2021

Major: Civil Engineering

