IMAGE ANALYSIS OF UNSUPERVISED CLASSIFICATION IN LAND COVER TYPES BY GIS TECHNIQUE: SOUTHERN YANGON REGION

Win Thet Myint[1]

ABSTRACT

This research approach image classification using Arc GIS from satellite Landsat image. Supervised Learning predicts based on a class type. Unsupervised Learning discovers underlying patterns. Image classification is small piece of the very intricate machine learning. In addition to classification, other key aspects include object detection and object localization. This is followed by the extended methodological workflow supported by illustrative print screens and technical description of data processing in ArcGIS. The methodology includes a workflow involving several technical steps of raster data processing in ArcGIS: 1. coordinate projecting, 2. pre-processing, 3. inspection of raster statistics, 4. spectral bands combination, 5. calculations, 6. unsupervised classification, 7. mapping. The classification was done by clustering technique using ISO Cluster algorithm Classification. The main purposes of this research are to interpret changing the area of the land use land cover in the interest study area using the Unsupervised Classification method, Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Built Up Index (NDBI) approaches. This paper finally presents the results of the ISO Cluster application for Landsat (8) image processing and concludes final remarks on the perspectives of environmental mapping based on Landsat (8) image processing in ArcGIS.

Keywards: Supervised Classification, Unsupervised Classification, NDVI, NDBI

[1] Lecturer, Department of Economic Geography, Co-operative University, Thanlyin

1.INTRODUCTION

Image classification refers to the process of extraction of information from satellite images. The purpose of image classification is to categorize all pixels in a digital image into one of several land use classes or themes. In order to make use of the multitude of digital data available from satellite imagery, it must be processed in a manner that is suitable for the researcher.

The interest study area is located within Yangon and a small part of the Bago Region. It comprises built-up areas, waterbodies, vegetation, cultivated land, etc. The main focus of my research is calculating the image analysis of land use and land cover patterns using Arc GIS software. The paper presents the use of the Landsat 8 image processed by the Spatial Analyst Tool for raster data processing, band calculations, and classification. Landsat 8 imagery is currently one of the most widely used satellite-based earth observation programs.

For my research, this processing includes categorizing the land into its various use functions. Depending on the interaction between the computer and the interpreter during the classification process, there are two types of classification. These two main categories used to achieve classified output are called supervised and unsupervised classification techniques. Unsupervised classification (clustering) is an effective method of partitioning remote sensor image data in multispectral feature space and extracting land-cover information.

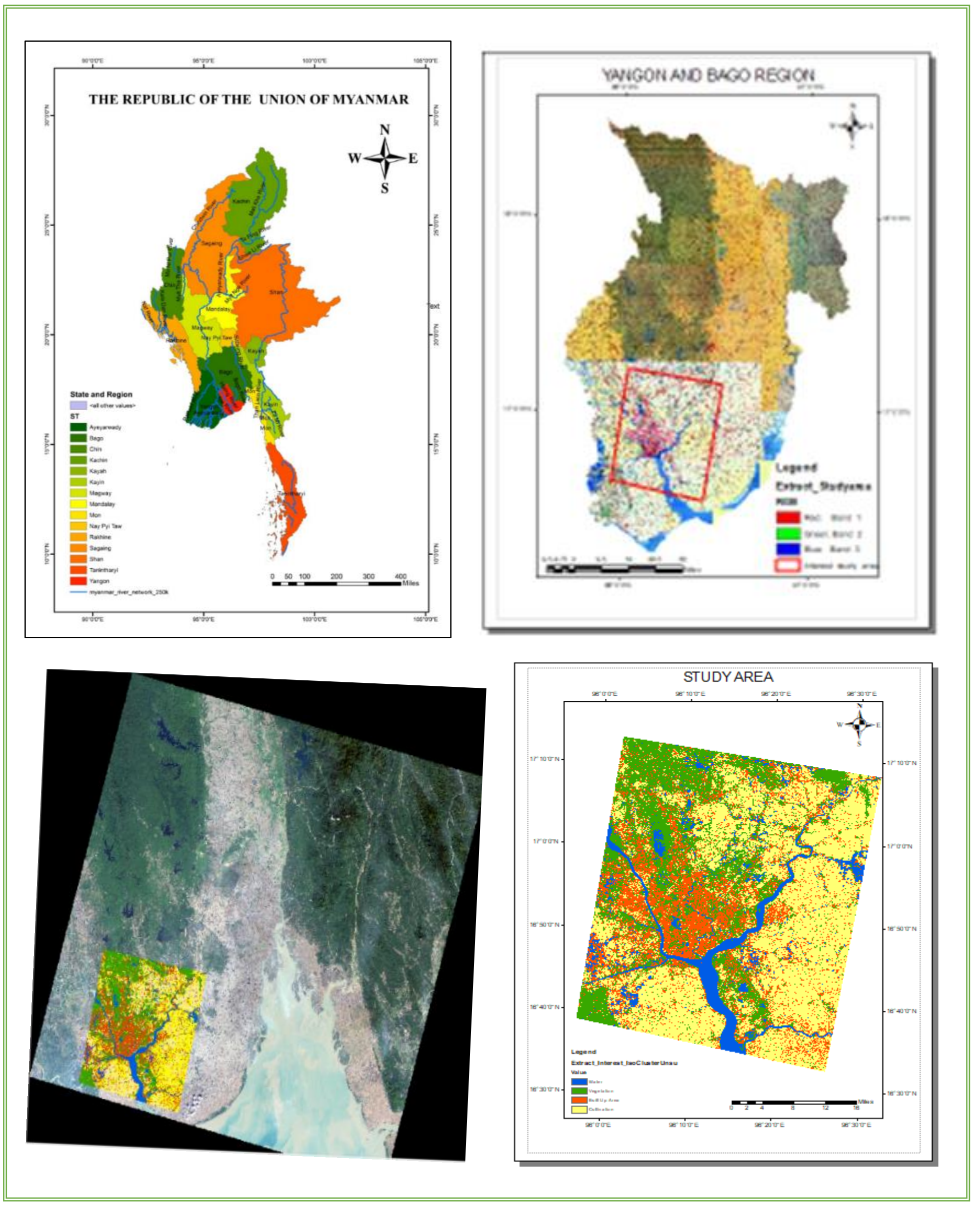
Because of incessant urbanization, land cover in urban areas tends to change more drastically over a short period of time than elsewhere. Urbanization has led land cover to change, especially frequently in sub-urban areas in Yangon Region as a result of rapid economic development. It is based on the combinational use of NDBI and NDVI. The mapping is accomplished through arithmetic manipulations and recoding of NDBI and NDVI images derived from a 2022 Landsat image.

This method also produces very accurate results, more efficiently than the unsupervised classification method. Research has proved a significant rapid land cover change and a vast transformation of agricultural and forest land into low-density urbanized areas scattered in the study area. A rapid change in land cover in a regional development corridor has a significant impact on urban expansion, particularly in the periphery.

II. INTEREST OF STUDY AREA

The interest study area is largely the urban area of Yangon City and its surrounding regions. It is located in the southern part of the Yangon Region and lies between 16° 30' 0" North latitude and 17° 10' 0" North latitude and 96° 00' 30" East longitude and 96° 30' 0" east longitude. The study area covered (3501.4701) sq.km and Yangon is the region that is densely populated with urban areas.

Figure (1) INTEREST OF STUDY AREA



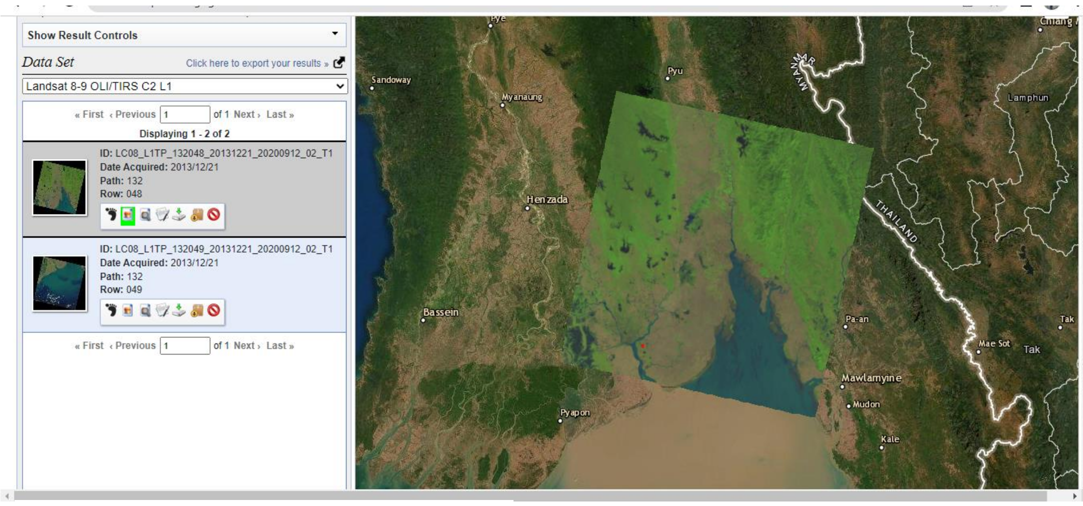
Source: Myanmar Information Management Unit, Yangon

III.MATERIALS AND METHODOLOGY

3.1 Materials Data and Software Used

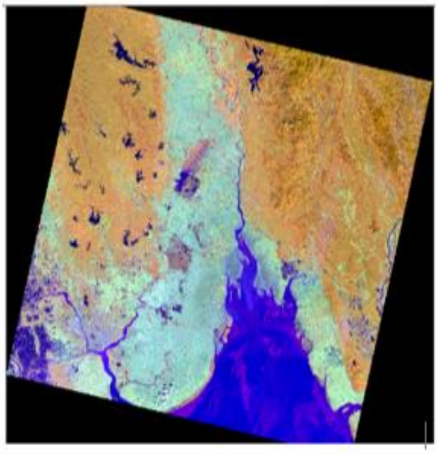
In this research paper, satellite image data is needed for classification of land use and land cover mapping and is associated with NDVI and NDBI of the study area of interest. The data was downloaded for free from the United States Geological Survey (USGS) and the resolution was 30 meters. All the data was processed using Arc GIS software. Landsat (8) records data in eleven different bandwidths. These bandwidths cover parts of the visible, infrared, and thermal infrared regions of the electromagnetic spectrum.

Figure (2) Characteristics of Landsat (8) and Image



Source: Earth Explorer (USGS)

Figure (3) Characteristics of Landsat (8) and Image



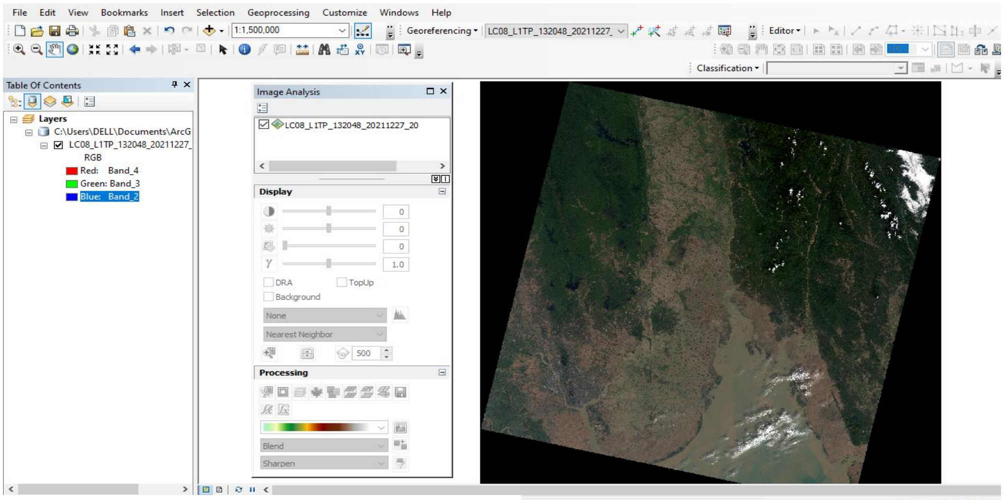
	Bands	Wavelength (micrometers)	Resolution (meters)
Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Launched February 11, 2013	Band 1 - Coastal aerosol	0.43 - 0.45	30
	Band 2 - Blue	0.45 - 0.51	30
	Band 3 - Green	0.53 - 0.59	30
	Band 4 - Red	0.64 - 0.67	30
	Band 5 - Near Infrared (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: Earth Explorer (USGS)

3.2 Image Pre-Processing

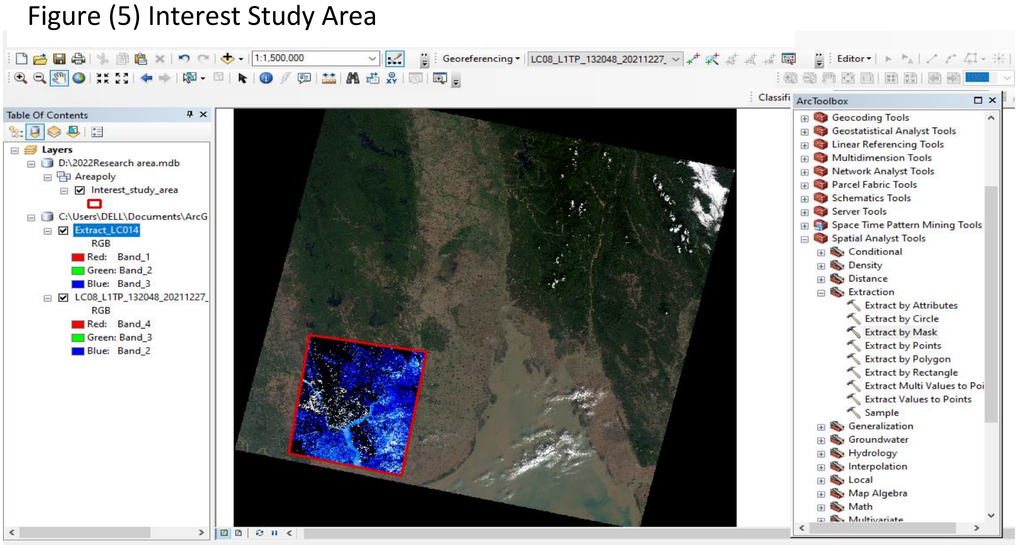
The image classification analysis is based on the unsupervised methods in this paper. Before interpreting land use land cover (LULC), all bands of satellite images will be georeferenced on the UTM WGS 1984 coordinate system (Zone 47). The clustering process results in a classification map consisting of spectral classes. The analyst then attempts to transform the spectral classes into thematic information classes of interest (e.g., urban, water, agriculture). where the researcher has to make decisions on which categories can be grouped together into a single land use category.

Figure (4) Pre-Processing of Satellite Image



Source: Researcher Analysis

Figure (5) Interest Study Area



Source: Researcher Calculation

Moreover, the bands will be cropped and converted from the digital number into the reflectance value of each band. Equations (1) and (2) are used to convert the digital number (DN) to a reflectance value.

Top of Atmosphere (TOA) Radiance

$$L\lambda = ML * Qcal + Al$$

1. $L\lambda$ = TOA spectral radiance (Watts/ (m2 * sr * µm))
2. ML = Radiance multiplicative Band (No.)
3. AL = Radiance Add Band (No.)
4. Qcal = Quantized and calibrated standard product pixel values (DN)

(2) Correct from sun Azimuth

$$\text{Correct from sun Azimuth} = TOA / \Theta \text{ (Sun Elevation)}$$

(3) Normalized Differential Vegetation Index (NDVI)

$$NDVI = (NIR - RED) / (NIR + RED)$$

1. RED (Band 4) = DN values from the RED Band
2. NIR (Band 5) = DN values from Near-Infrared Band

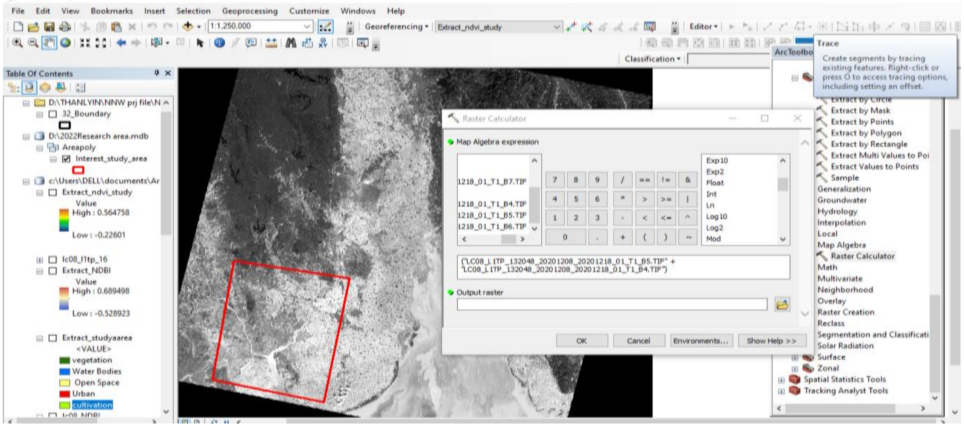
(4) Normalized Difference Built-up Index (NDBI)

$$NDBI = (SWIR - NIR) / (SWIR + NIR)$$

1. SWIR (Band 6) = DN value from Short Wave INFRARED
2. NIR (Band 5) = DN values from Near-Infrared Band

Land use classification and land cover are done by the unsupervised classification method on the value of NDVI and NDBI. To validate land use and land cover, the interpretation results of Google Earth image and UTM (1:50000) and NDVI & NDBI values is used. Identification of land use and land cover is based on the results obtained from NDVI and NDBI reflectance values.

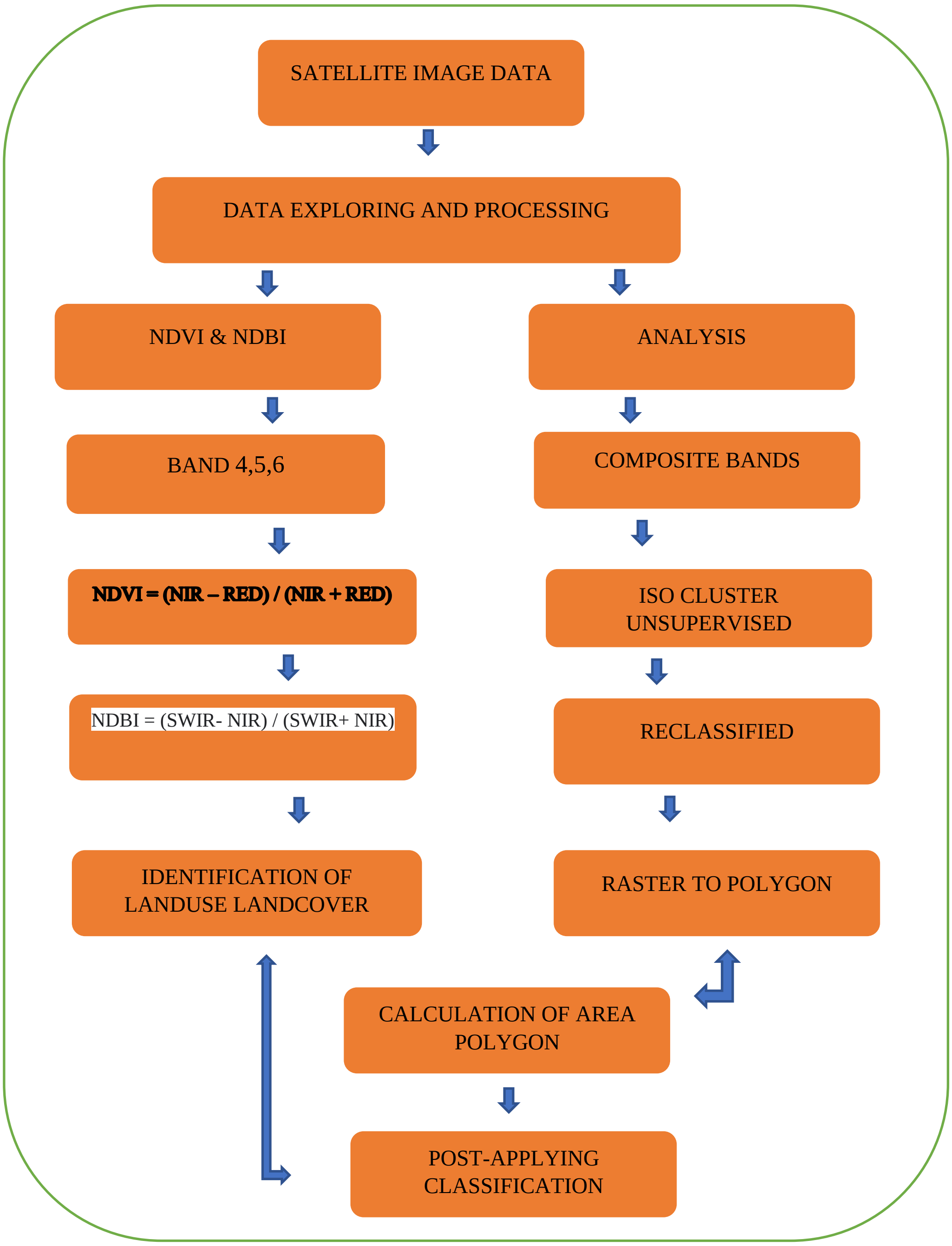
Figure (6) Calculation for NDVI and NDBI



Source: Researcher Calculation

3.3 Work Flow

Figure (7) Process of Work Flow



Source: Researcher Analysis

The secondary data in the work flow diagram is a satellite image of Landsat (8) data, after which the research paper analysis is calculated step by step using the Arc GIS. Finally, the result of land use and land cover patterns of interest in the study area is shown by the area polygon on the percentage of various land use patterns.

IV. RESULT AND DISCUSSION

4.1 Image Processing

In image analysis, unsupervised image classification and normalized difference built-up index (NDBI) were used to analyze the land use pattern in the study area. Unsupervised image classification is based entirely on the automatic identification and assignment of image pixels to spectral groupings. It considers only spectral distance measures and involves minimal user interaction. The normalized difference The National Built-up Index (NDBI) uses the NIR and SWIR bands to emphasize manufactured built-up areas

The unsupervised classification and Normalized Difference Built-up Index (NDBI) were made using ArcGIS, (30) meter resolution Landsat (8) satellite image to classify and distinguish where cultivated habitats, urban areas, water bodies, and vegetation areas can be separated into layers of several classes. The research focused on the region bordered by 16° 30' 0'' North latitude to 17° 10' 0'' North latitude and 96° 00' 30'' East longitude to 96° 30' 0'' East longitude and urban built-up areas with various land cover types.

4.1.1 Interpretation of color compositions

Unsupervised classification is where you let the computer decide which classes are present in an image based on statistical differences in the spectral characteristics of pixels. After the unsupervised classification is complete, we need to assign the resulting classes into the class categories within the schema.

The satellite image of Landsat (8) was processed and color composites were created using seven band combinations in this research paper. A variety of band combinations provide additional information on what can be highlighted in the correct colors. The unsupervised classification based on the color composites does not require creating training points and is a fully machine learning approach. The clustering algorithms here were used to build training sites without using field survey data.

When interpreting the results of the land cover classification, it can be discovered that classifying spectral reflectance of pixels by machine learning made fewer errors in comprehending land cover types and distinguishing between landscape classes than it did with human-based GIS classification.

1. The combination of natural colors is achieved by the bands of (3-2-1).
2. The natural color composite uses a band combination of red (4), green (3), and blue (2). While healthy vegetation is green, unhealthy flora is brown. Urban features appear white and grey, and the water is dark blue or black.
3. It uses near-infrared (5), red (4), and green (3). This band composition is useful for analyzing vegetation. In particular, areas in red have better vegetation health. Dark areas are water, and urban areas are white.
4. The short-wave infrared band combination uses SWIR-2 (7), SWIR-1 (6), and red (4). These composites display vegetation in shades of green; urban built-up areas are blue; and soils are brown.
5. This band combination uses SWIR-1 (6), near-infrared (5), and blue (2). It's commonly used for crop monitoring because of its use of short-wave and near-infrared. Healthy vegetation appears dark green. But bare earth has a magenta hue.
6. The geology band combination uses SWIR-2 (7), SWIR-1 (6), and blue (2). This band combination is particularly useful for identifying geological formations, lithological features, and faults.
7. The bathymetric band combination (4,3,1) uses the red (4), green (3), and coastal band to peak into water. The coastal band is useful in coastal, bathymetric, and aerosol studies because it reflects blues and violets. This band combination is good for estimating suspended sediment in the water.

4.2 Image Analysis

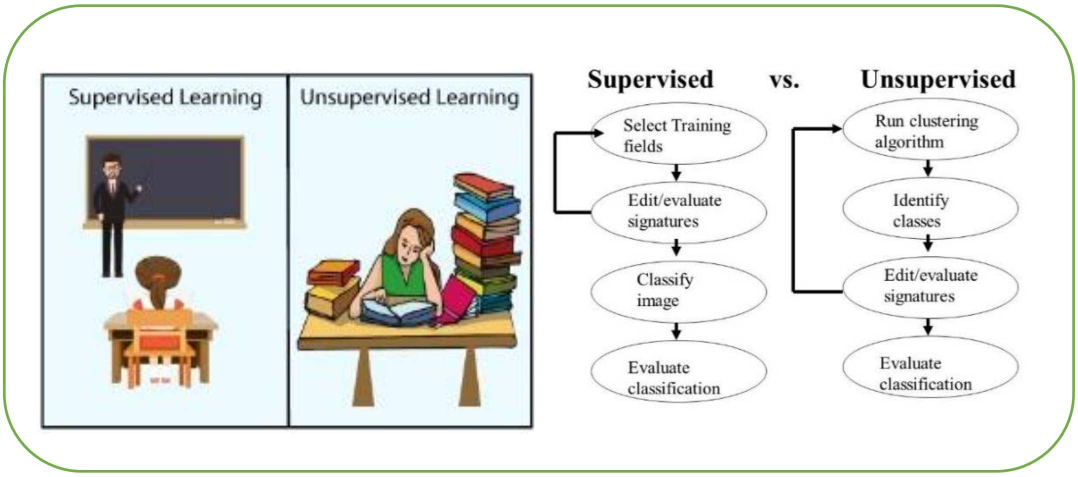
4.2.1 Unsupervised Image Classification

There are two major categories of image classification techniques: **unsupervised** (calculated by software) and **supervised** (human-guided) classification.

Supervised classification is based on the idea. Researcher can select sample pixels in an image that are representative of specific classes and then direct the image processing software to use these training sites as references for the classification of all other pixels in the image. Training sites (also known as testing sets or input classes) are selected based on the knowledge of the researcher.

Unsupervised classification is where the researcher bases their analysis on the software analysis of an image without providing sample classes. Classification is done using one of several statistical routines generally called "clustering," where classes of pixels are created based on the spectral signatures. The researcher can specify which algorithm the software will use and the desired number of output classes, but otherwise does not aid in the classification process.

Figure (8) Supervised and Unsupervised Classifrication



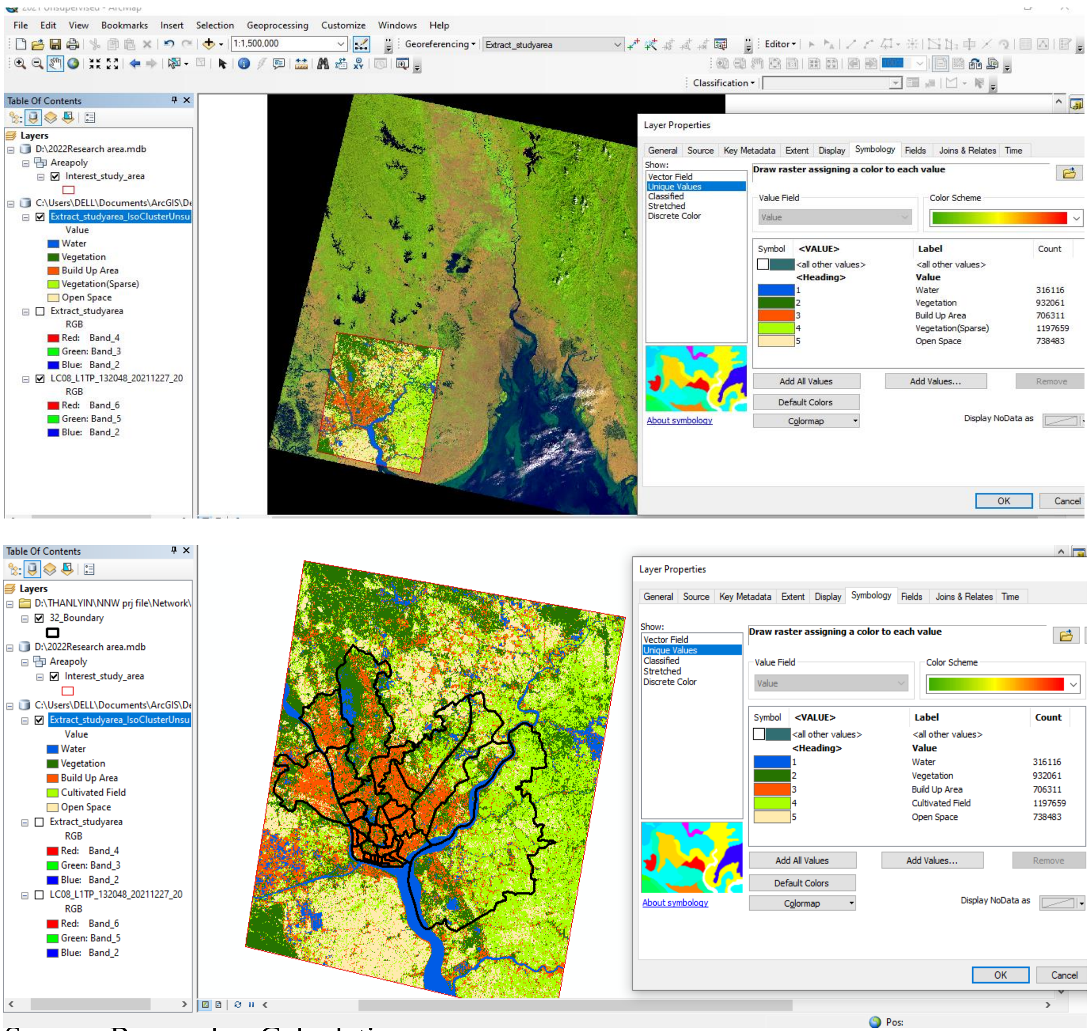
Source: Google Image

However, the researcher must have knowledge of the area being classified when the groupings of pixels with common characteristics produced by the computer have to be related to actual features on the ground (such as wetlands, developed areas, forests, etc.). It is up to the analyst to assign meaningful names to the classes after processing.

Iso Cluster Unsupervised Image Classification is the only unsupervised classification tool provided. Unsupervised classification (commonly referred to as clustering) is an effective method of partitioning remote sensor image data in multispectral feature space and extracting land-cover information. "***Unlike supervised classification techniques, unsupervised classification does not require the user to generate a training set prior to classification and is thus less time-intensive.***" *(Albert, 2002).*

In the process of unsupervised image classification, Band No. 1 to Band No. 7 are first selected from the Landsat (8) image. These bands are Image Analysis > Composite Band and then extract the interesting parts of the study area. After that, it was calculated as an ISO Cluster from the extracted area of a raster image and reclassified into five categories.

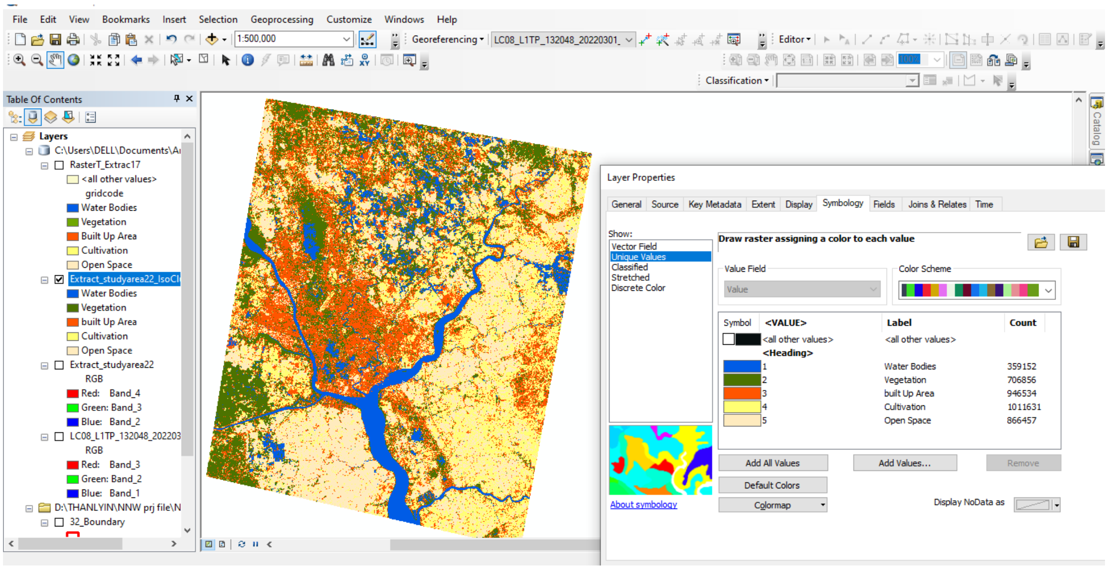
Figure (9) ISO Cluster Unsupervised Classification in Study Area



Source: Researcher Calculation

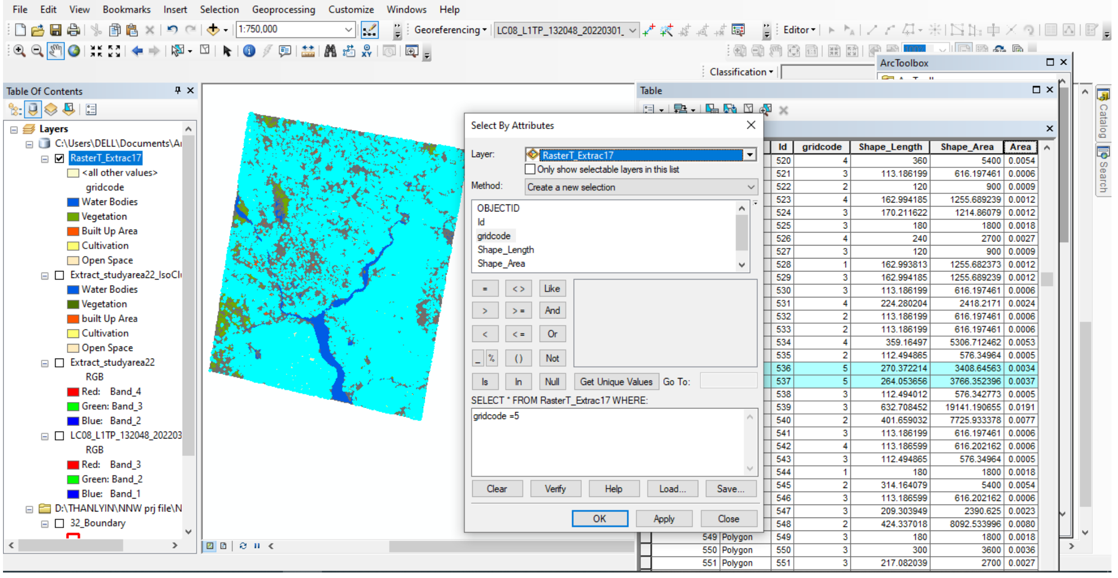
The Iso Cluster Unsupervised Classification tool automatically finds the clusters in an image and outputs a classified image. It is based on the Iso Cluster tool. These land use patterns are: water bodies, built-up areas, vegetation (forest/trees), cultivation, and open space. To compute each polygon area, calculate the reclassified aster image in a polygon and find the area in the Attribute Table from ArcGIS software.

Figure (10) ISO Cluster Unsupervised Classification in Study Area



Source: Researcher Calculation

Figure (11) Compute the polygon Area with Geocode



Source: Researcher Calculation

Table (1) Major Attributes of Land Cover Classification in Study Area

No.	Land Cover	Major Attributes
1	Built Up Area	Residential, commercial, industrial, transportation, built up rea
2	Vegetation	agricultural land and forest
3	Water Bodies	Rivers, streams, lakes, drains, and Dams
4	Cultivation	Paddy field, Crop land
5	Open Space	Urban parks, shrubs, barren land

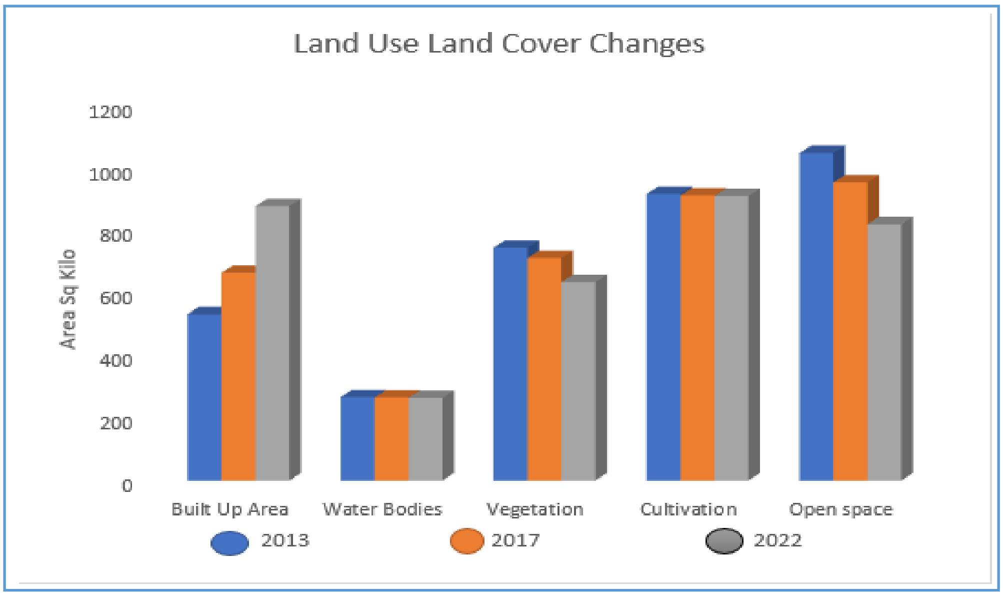
Source: Researcher Calculation

Table (2) Variation of Land use Landcover Area Changes in Study Area (2013-2022)

No	Categories	2013 (Area Sq kilo)	2017 (Area Sq kilo)	2022 (Area Sq kilo)	Changes (2013-022) (Area Sq kilometer)
1	Built Up Area	531.10	663.23	876.45	345.55
2	Water Bodies	266.07	265.21	264.01	-2.06
3	Vegetation	743.20	710.55	633.90	-109.3
4	Cultivation	915.20	910.23	908.58	-6.62
5	Open space	1045.90	952.25	818.53	-227.37
	Total	3501.47	3501.47	3501.47	_

Source: Researcher Calculation

Figure (12) Changes of Land Use Land Cover in Study Area



Source: Researcher Calculation

Five class categories are divided with area polygon features to show the different land cover classifications in the study area. According to the results of unsupervised classification, the built-up area covers approximately (531.10) square miles (15.16%) in 2013 and approximately (876.45) square miles (25.03%) in 2022, implying that the land cover area has increased by (345.55) square kilometers (25.03%) in (10) years.

Research has shown that with the significant increase in urban built-up areas, vegetation and cultivation areas are declining. Cultivated areas close to the sub-urban area can be included in the urban area due to urban expansion. In the absence of field measurements, Arc GIS Software can be used for analysis using satellite imagery and to conduct regional research.

4.2.2 NDVI (**(Normalized Difference Vegetation Index)**

Land use classification and land cover are done by supervised classification method on the value of NDVI and NDBI. Ground check to the field is carried out to obtain the validation of land use and land cover from interpretation result of NDVI and NDBI value. Identification of land use and land cover are based on the result obtained from NDVI and NDBI reflectance value.

The NDVI approach is the simplest way to identify the level of vegetation density in the area. There is the specific wavelength used in the NDVI approach in which the red band (band 4) and the near-infrared (band 5). The presence of near-infrared spectra (NIR) and red spectrum (RED) in Landsat 8 make the identified easier to determine the vegetation on the land cover by using vegetation index (NDVI).

NDVI has a value range from (-1) to (+1) and the value of NDVI 0 indicates the non-vegetated area. The higher the NDVI values, the greater the density of vegetation. Maximum reflection at near-infrared wavelengths and absorption of light in the red will occur in the vegetated area. In the vegetated area, maximum reflection and absorption of light in the red will occur. The higher the value of NDVI reflects high near-infrared (NIR), which means dense greenery. Generally, we obtain the following result:

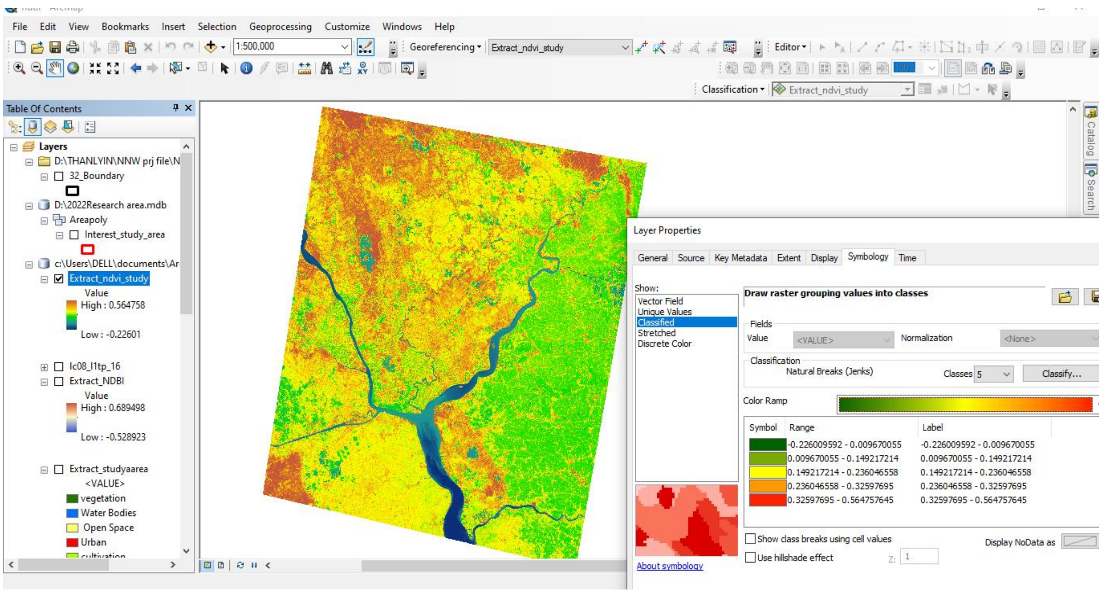
- Water bodies are represented by NDVI values ranging from (-1) to (0).
- NDVI values ranging from (-0.1) to (0.1) represent barren rocks, sand, or snow.
- NDVI = (0.2) to (0.5) denotes shrubs and grasslands, as well as senescing crops.
- NDVI values ranging from (0.6) to (1.0) indicate dense vegetation or tropical rainforest.

The NDVI rate can be calculated using the raster calculator in ArcGIS.

The maximum reflection will occur in near-infrared waves caused by leaf structure (mesophyll), which increases the reflection of near-infrared, while at the red wavelength there will be maximum absorption caused by green leaf substance (chlorophyll) "*(Molidena and Assyakur, 2012).* The results show that the study area has (-0.2260) and (0.56476) as the minimum and maximum values of NDVI, respectively.

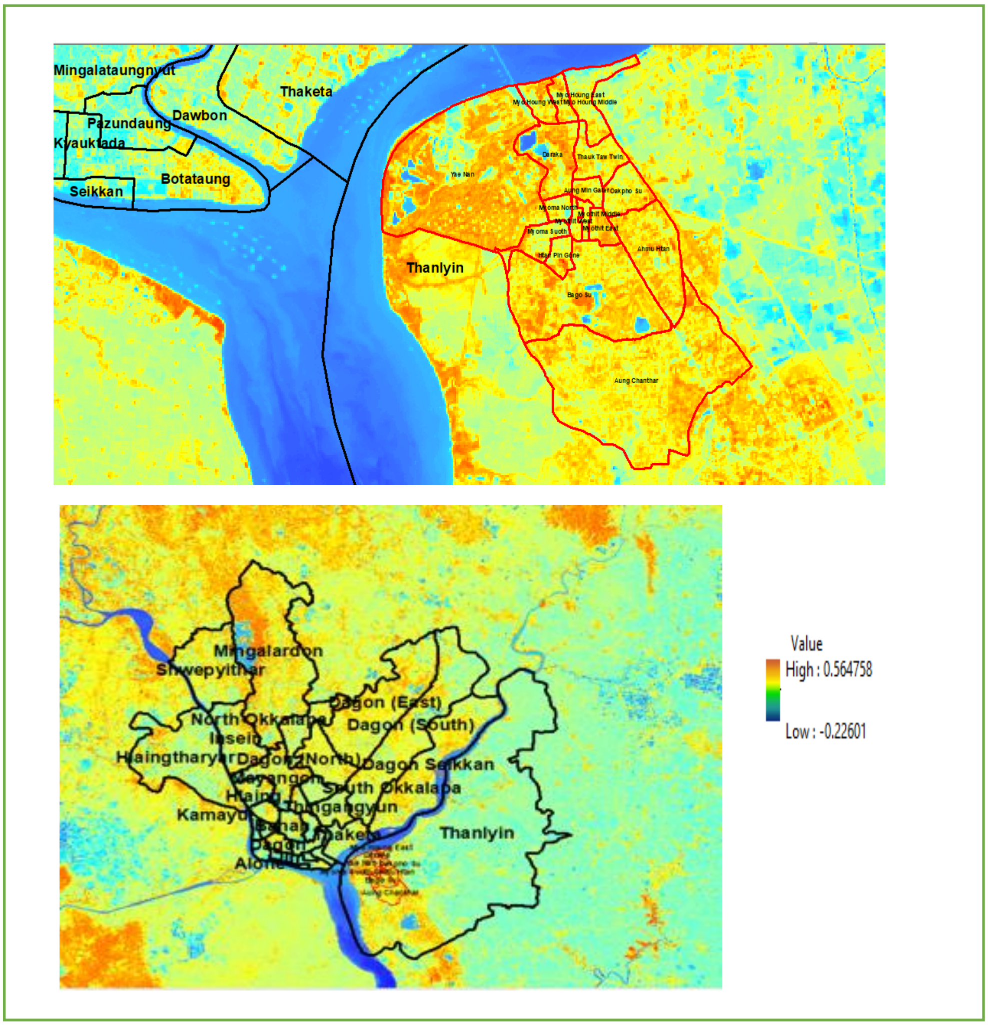
NDVI values which are greater than 0.500 indicate densely vegetated areas, whereas the negative NDVI values indicate non-vegetated areas such as water and snow. NDVI values between 0.002 and 0.025 indicate open land and settlements. According to "*Sudiana* and *Diasmara*" (2008), the area that has NDVI values smaller than 0.2 is a territorial water or rocky soil, and the NDVI values greater than 0.4 indicate vegetated and fertile vegetation.

Figure (12) Calculation of NDVI Value in Study Area



Sources: Researcher Calculation

Figure (13) Calculation of NDVI for Thirty-Two Townships and Thanlyian



Source: Researcher Calculation

4.2.3 NDBI (Normalized Difference Built-up Index)

Image classification techniques (supervised classification and unsupervised classification) are lengthy and complex processes. It requires a composite band and the application of a number of operations for the final result. The accuracy derived from the image classification technique depends on the image analyst and the method followed by the analyst. However, the NDBI calculation is simple and easy to derive.

For Landsat 8 data, NDBI = (Band 6 – Band 5) / (Band 6 + Band 5)

The normalized difference The National Built-up Index (NDBI) uses the NIR and SWIR bands to emphasize manufactured built-up areas. The SWIR spectrum (band 6) can detect geological, soil, and rock conditions while the near-infrared spectrum (band 5) is used to identify the biomass/vegetation content.

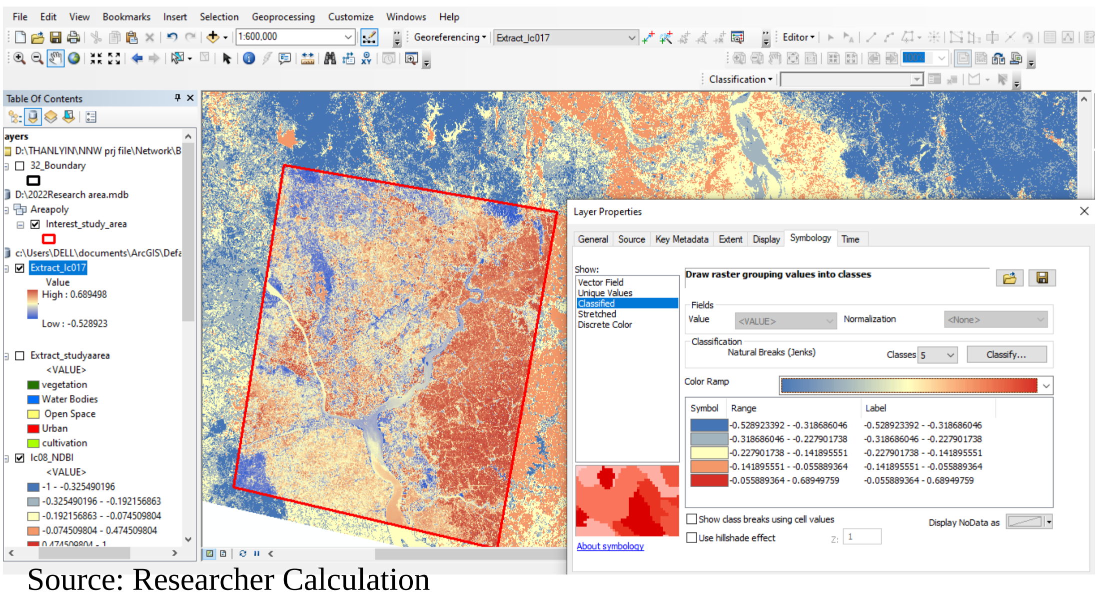
Also, the normalized difference build-up index value lies between (-1) and (+1). The negative value of NDBI represents water bodies, whereas the higher value represents built-up areas. The NDBI value for vegetation is low. *"Positive values of NDBI represent urban land areas and negative values of NDBI represent non-urban land areas" (Zha et al., 2003).*

The negative NDBI shows the unopened area (or) vegetation area, whereas the positive NDBI indicates the open area. When the SWIR spectrum concerns the settled object, the reflected reflectance value will be large, which means when it comes to the plant as the object, the reflectance value indicated by the small SWIR spectrum will be small. The developed area has a greater reflectance value on band (6) compared to band (5). Band (6) will reflect maximally when reflected by settlements or open areas.

"Zha et al. (2003) mentioned that the NDBI method was the most effective and profitable to map the built area so that it could be an alternative to mapping the rapid development of the city." The NDBI method cannot separate the types of areas such as settlements, industrial estates, and other commercial areas. In addition, NDBI cannot separate the open ground because it has the same spectral response in the SWIR and NIR spectra. To distinguish buildings from bare or empty areas, open waterbodies, and vegetation can be done by extracting the built-up area with MNDWI. According to the result of the analysis, the value of NDBI in the study area is (-0.5289) up to (0.6895).

The dense settlements and developed areas are mostly located in the study area around Yangon City, near coastal areas, and the vegetation is mostly located in the northern part of the study area.

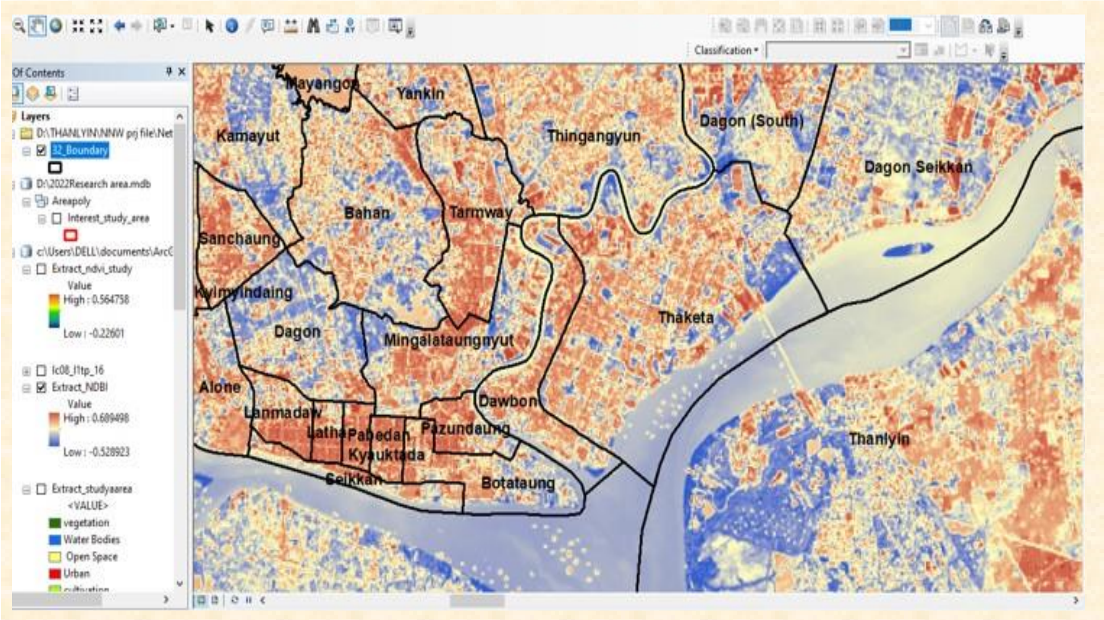
Figure (13) Normalized Difference Built-up Index (NDBI)in Study Area



Source: Researcher Calculation

There are many indexes for the analysis of Urban built-up areas. Normalized Difference the Built-up Index (NDBI), Built-up Index (BU), Urban Index (UI), Index-based Built-up Index (IBI), Enhanced Built-up and Bareness Index (EBBI) are the most common indexes for analysis of the built-up areas.

Figure (14) Overlay mapping of NDBI Value and Townships Area



Source: Researcher Calculation

These different indexes have their own formula and their own calculation method. The bare soil and built-up areas reflect more SWIR than NIR. The water body does not reflect in the infrared spectrum. -up index is the binary image with only the highest positive value indicating built-up and barren areas. This allows BU to map the built-up area automatically.

BU = NDBI-NDVI = Image classification techniques (supervised classification and unsupervised classification) are lengthy and complex processes. It requires a composite band and the application of a number of operations for the final result. The accuracy derived from image classification technique depends on the image analyst and the method followed by the analyst.

However, the NDBI calculation is simple and easy to derive. The negative value of NDBI represents water bodies, whereas the higher value represents built-up areas. The NDBI value is low in vegetation.

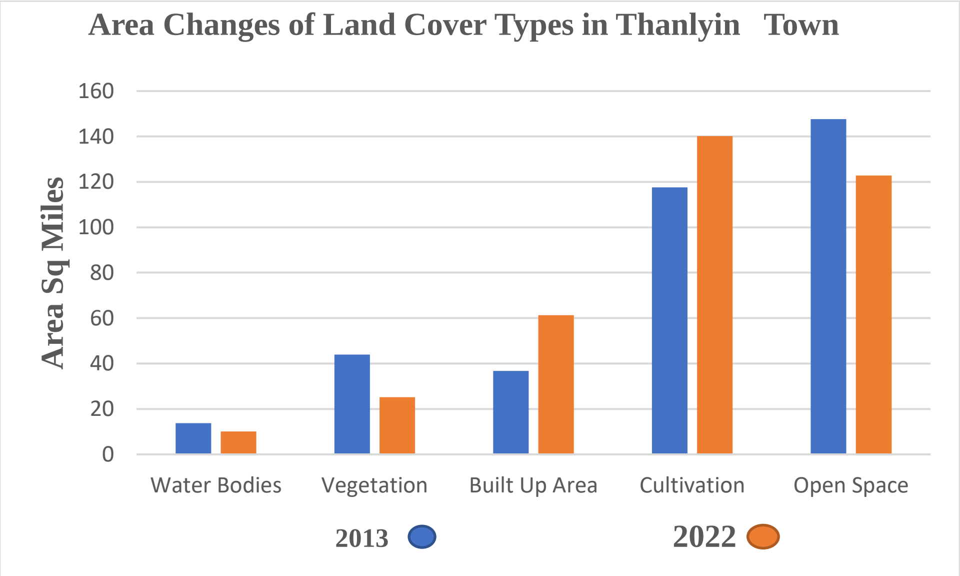
The value of NDBI ranges between (−1) and (1). Research suggests that positive values of NDBI represent urban land areas and negative values of NDBI represent non-urban land areas. In general, using NDVI and NDBI approaches, it is seen that the study area mostly consists of vegetation area and a small amount of built area. Land use/land cover in the interest study area is dominated by forest, trees, shrubs, rice fields, settlements, bare soils and other.

Table (3) Variation of Land use Landcover Area Changes in Thanlyin Township

No	Categories	2013 (Area Sq Kilo)	2022 (Area Sq Kilo)	Changes of Area (2013-2022)
1	Water Bodies	13.71	10.06	13.65
2	Vegetation	43.98	25.14	18.84
3	Built Up Area	36.70	61.33	24.63
4	Cultivation	117.63	140.17	22.54
5	Open Space	147.55	122.87	24.68
	Total	137.59	137.59	_

Source: Researcher Calculation

Figure (13) Changes of Land Use Land Cover in Thanlyin Township



Source: Researcher Calculation

V. CONCLUSION

The satellite observations can provide globally consistent and repetitive measurements of the Earth's surface conditions, returning important data with different spatial, spectral, radiometric, and temporal resolutions. They are very useful for monitoring the spatial distribution and growth of urban built-up areas because of their ability to provide timely and synoptic views of land cover.

Land use changes in urban areas usually occur because of high urbanization and residential development rates, causing the suburbanization process and consequently urban sprawl (generally known as uncontrolled urban development), which implies negative effects on human communities, such as overcrowding, traffic congestion, open space loss, increasing pollution, poor housing quality, etc.

Recent technological advances in Geographic Information Systems (GIS) and image analysis offer the potential of using digital images to objectively quantify ground cover and the composition of vegetation in a repeatable and timely manner. Land use in urban areas is very dynamic and changes continuously, mainly due to the construction of new buildings, roads, and other infrastructure correlated with ongoing natural land use.

Unsupervised classification methods govern spectral classes automatically but generally show a restricted capacity to precisely partition the landscape into normal classes. Unsupervised classification depends on the purpose of study and the condition of the selected area. This study assessed the effectiveness of using the Landsat (8) image for the automatic classification of the land cover types in the study area using ArcGIS. The automation in Earth observation studies is based on using machine algorithms for data recognition and processing.

The combination of the satellite spectral data from Landsat (8) imagery in ArcGIS demonstrated classification of the scenes for environmental monitoring and land policy analysis in the study area of Yangon, Southern Region. (e.g., built-up areas, water bodies, agricultural crop fields). Although the visible bands of the Landsat (8) have a moderate accuracy (30 m), to increase the accuracy and resolution of the image of unsupervised classification, as demonstrated.

An ISO Cluster approach to classifying the image composed of the three spectral bands is a promising approach for land cover type mapping. This study presented mapping land cover types using ArcGIS by the unsupervised classification approach. The methodology includes several print screens illustrating the workflow, which can be repeated in similar research in the future.

References

Andriani, D. D. (2018). Interpretation of land use and land cover at lowland area using by NDVI and NDBI . *Eco. Env. & Cons*, 24(2).

Lemenkova, P. (2021). ISO CLUSTER CLASSIFIER BY ARCGIS FOR UNSUPERVISED CLASSIFICATION OF THE LANDSAT TM IMAGE OF REYKJAVÍK. *Bulletin of Natural Sciences Research*, 29-37.

LEMENKOVA, P. (2021). ISO CLUSTER CLASSIFIER BY ARCGIS FOR UNSUPERVISED CLASSIFICATION OF THE LANDSAT TM IMAGE OF REYKJAVÍK . *Bulletin of Natural Sciences Research*.

Martins A. Oyekola, G. K. (2018). Unsupervised Classification in Land Cover Types Using Remote Sensing and GIS Techniques . *International Journal of Science and Engineering Investigations* .

Muhammad Yazrin Yasin, J. A. (2020). Land Cover and NDBI analysis to map built up area in Iskandar Malaysia . *IOP Conf. Series: Earth and Environmental Science* .

Muhammad Yazrin Yasin, J. A. (2020). Land Cover and NDBI analysis to map built up area in Iskandar Malaysia . *Earth and Environmental Science*.