A STUDY ON LAND COVER EXTRACTION AND MAPPING IN CAGAYAN DE ORO RIVER BASIN USING GOOGLE EARTH ENGINE

Stephanie Mae Salcedo-Albores*[1,2], Kristine Sanchez-Companion[2] and Elizabeth Edan M. Albiento[2]
[1]Department of Civil Engineering, College of Engineering & Geosciences, Caraga State University
Ampayon, Butuan City 8600
Email: ssalbores@carsu.edu.ph

[2]Department of Civil Engineering, MSU-Iligan Institute of Technology
Tibanga, Iligan City 9200
Email: kristine.sanchez@g.msuiit.edu.ph, elizabethedan.albiento@g.msuiit.edu.ph

KEY WORDS: Google Earth Engine, Land Cover, Remote Sensing, River Basin

ABSTRACT: Land cover is a crucial aspect in remote sensing analyses, especially for land cover classification using remotely sensed imagery. Remote Sensing, coupled with the spatial data analysis capabilities of GIS, has proven effective in characterizing land cover change patterns, intensity, and dynamics over large areas accurately and promptly. Google Earth Engine, an online platform, offers a catalog of satellite images and geospatial datasets for Earth science data and analysis. In this study, custom scripts were developed, incorporating various components from official Google resources and other references, to extract and process land cover maps from Google Earth Engine's satellite images. The 2021 Land Cover Classification achieved a user's accuracy of 95.01% and a Kappa index of 94.31%. User's accuracy represents false positives, where pixels are misclassified as a known class when they should have been classified differently, known as errors of commission. On the other hand, the Kappa index of agreement provides an overall assessment of the classification's accuracy. The dominant land cover class in 2021 was the Closed Forest, mainly concentrated in the middle and upstream regions of the basin. The Open Forest class covered smaller scattered areas in the middle and upstream parts. Grassland and agricultural land ranked third and fourth, respectively. To assess the accuracy of the classified map, ArcMap was utilized, employing the common method of creating random ground truth points and comparing them with the classified data using a confusion matrix. One thousand five hundred forty-three (1,543) random ground truth points were generated using the create accuracy assessment points tool for the 2021 Land Cover classification. The Compute Confusion Matrix Tool was then used to calculate the accuracy of the classification. These tools ensure that each point has valid class values for both the classified and ground truth data. The confusion matrix provides insights into errors of omission and commission, from which the kappa index of agreement and overall accuracy between the classified map and reference data are derived. Overall, this study demonstrates the effectiveness of Google Earth Engine in extracting and mapping land cover in the Cagayan de Oro River Basin. The high accuracy of the classification showcases its potential for large-scale land cover analyses, which is essential for environmental monitoring and resource management.

1. INTRODUCTION

1.1 Rationale

Land cover is a crucial aspect in remote sensing analyses, especially for land cover classification using remotely sensed imagery. Remote Sensing, coupled with the spatial data analysis capabilities of GIS, has proven effective in characterizing land cover change patterns, intensity, and dynamics over large areas accurately and promptly. Google Earth Engine, an online platform, offers a catalog of satellite images and geospatial datasets for Earth science data and analysis. In this study, custom scripts were developed, incorporating various components from official Google resources and other references, to extract and process land cover maps from Google Earth Engine's satellite images. Most remote sensing analyses include a consideration of land cover. A clear example is the use of remotely sensed imagery for simple land cover classification (Franklin and Wulder, 2002; Alvarez et al., 2003 as cited in Aplin, 2004). The reason land cover is so important in remote sensing is because of how remotely sensed data is collected (Aplin, 2004). The capability of Remote Sensing to provide the most fundamental information describing the nature and extent of land cover, especially over large areas in an accurate and timely manner, together with GIS as an efficient tool for spatial data analysis, has been proven to be effective in characterizing the patterns, intensity and dynamics of land cover change (Petchprayoon et al., 2016, Santillan et al., 2011).

1.2 Objectives

The primary objective of this research is to extract and process land cover information from Google Earth Engine's satellite images with the aim of creating an accurate and up-to-date land cover map. This map will provide valuable insights into the distribution and composition of land cover types within the Cagayan de Oro River Basin. The study seeks to harness the capabilities of remote sensing and machine learning algorithms to achieve precise land cover classification, facilitating informed decision-making, environmental monitoring, urban planning, natural resource management, disaster response, and scientific research.

1.3 Study Area

The Cagayan de Oro River Basin was prepared and processed in ArcMap. Using the Synthetic Aperture Radar (SAR) – Digital Elevation Model (DEM), an estimate of 1,355.80 square kilometers of basin area was delineated. Figure 1 shows the location map of the study area, it traverses mostly to the Cagayan de Oro City and municipalities of Libona, Baungon and Talakag.

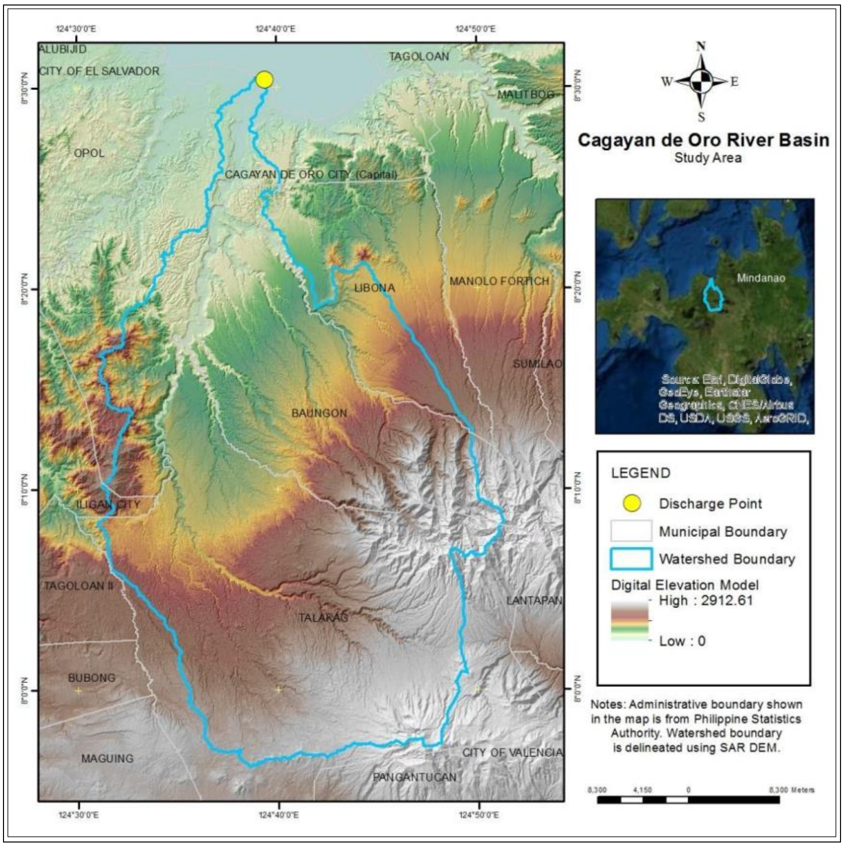

Figure 1. Location Map of Cagayan de Oro River Basin.

2. REVIEW OF RELATED LITERATURE

The necessity for precise data regarding the location, size, and nature of Land Cover (LC) is paramount across a spectrum of applications. Therefore, the creation of an up-to-date and precise LC map encompassing the entire country of Iran is imperative, utilizing the latest advancements in remote sensing, machine learning, and big data processing techniques. Additionally, it's crucial to establish an efficient method for automatic LC mapping over different time periods, eliminating the need for acquiring additional ground truth data across the vast expanse of this nation. Remote sensing has proven to be highly advantageous for mapping various LC categories. This research endeavors to produce the initial comprehensive LC map of Iran at a spatial resolution of 10 meters. This endeavor involved the utilization of 11,994 Sentinel-1 and 2,889 Sentinel-2 scenes collected in 2017. The Google Earth Engine (GEE) platform was harnessed to manage the extensive computational demands inherent in processing large time-series datasets covering the expansive geographical scope of Iran (Ghorbanian, et al., 2020).

The findings of Phan et.al. underscore the significance of selecting the appropriate dataset within the Google Earth Engine (GEE) platform, especially in regions with complex environmental conditions like Mongolia. These methods not only substantially reduce data volume, streamlining the analysis process, but also deliver accuracy on par with that of time series data. Nonetheless, it is worth noting that the spatial consistency among the classification results is relatively low compared to the overall high accuracy. This emphasizes the critical role of dataset selection within the GEE platform, especially within regions characterized by challenging conditions like Mongolia, where factors like snow cover, cloud cover, and expansive landscapes introduce additional complexity to land cover classification (Phan, Kuch, & Lehnert, 2020).

3. METHODOLOGY

Land cover data can be extracted through the available satellite images from Google Earth Engine (GEE). It combines planetary-scale analysis capabilities with a multi-petabyte collection of geospatial datasets and satellite pictures. Earth Engine is used by scientists, researchers, and developers to identify changes, chart trends, and measure variations on the surface of the planet. While still free for use in education and research, Earth Engine is now available for commercial usage (Google, 2019). Figure 2 contains the simplified flow in deriving land cover maps from raw satellite images extracted from the GEE, it will start with harvesting massive satellite data to produce land covers. After the satellite images are downloaded, they will be processed through an algorithm code to classify the image per pixel. Thorough classification of land cover type and validation will then be done in ArcMap.

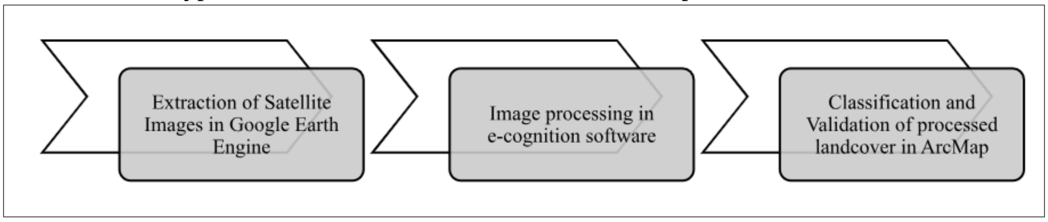

Figure 2. General flow for land cover processing and mapping.

The custom scripts developed in this thesis combine several components from official Google resources and other references. The methodology as shown in Figure 3 includes the following:

1. Defining the region of Interest (ROI).
2. Compositing the Landsat 8 image
3. Cloud Masking
4. Performing of OBIA classification – segmentation of the image to image-objects and classification of image-objects through SVM; and
5. Contextual editing of generated resource map; and
6. Accuracy assessments of generated base resource maps and refined resource maps.

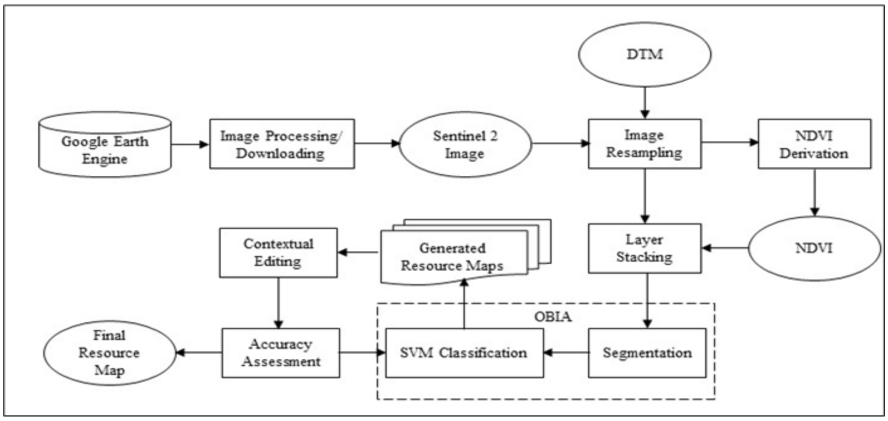

Figure 3. Conceptual Framework for land cover processing.

3.1 Image Processing

Study area boundaries in shapefile format were uploaded to extract just the area of interest - the Cagayan de Oro River Basin. This was done through the GEE's Interface. From importing the study asset, cloud masking, dataset of the imagery and the date of image. Sentinel visits the same spot on the Earth every 10 days. That means there will be around 36 images throughout the course of one year (and more where the scenes overlap). Google Earth Engine can be instructed to pick the median value in the stack. This has the benefit of removing clouds (which have a high value) and shadows (which have a low value). When using the median reducer to reduce an image collection, the composite value is the median in each band over time. Once the code is successfully executed, the image can now be exported into Google Drive.

3.2 Normalized Difference Vegetation Index (NDVI) Derivation

The NDVI is a basic yet effective index for assessing green vegetation that is well-known and widely utilized (Figure 4). It compares green leaf scattering in near-infrared wavelengths to chlorophyll absorption in red wavelengths to normalize green leaf scattering wavelengths. The NDVI has a value range of -1 to 1. Water is represented by negative NDVI values (values approaching -1). Generally, values near to zero (-0.1 to 0.1) correspond to barren rock or sand environments. Low, positive values (roughly 0.2 to 0.4) imply shrub and grassland, whereas high values indicate temperate and tropical rainforests (values approaching 1). It serves as a decent substitute for live green plants (Sahebjalal & Dashtekian, 2013).

$$NDVI = (NIR - Red)/(NIR + Red) \quad (1)$$

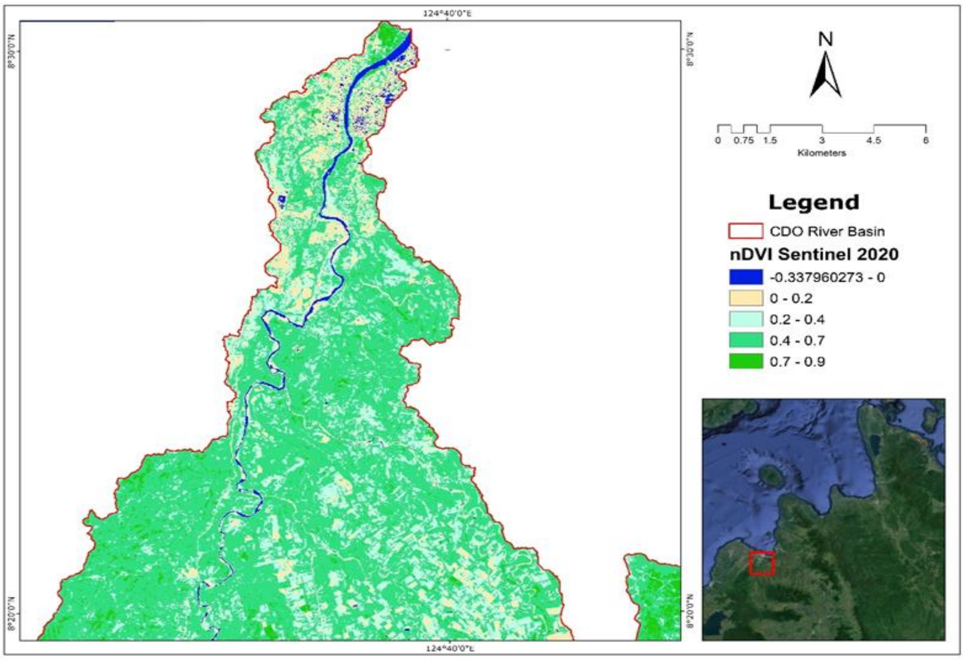

Figure 4. Sample Normalized Difference Vegetation Index (NDVI) year 2020.

3.3 Image Resampling and Layer Stacking

Because the bands of a Sentinel 2 image have varying resolutions, the band images are resampled to the highest resolution, which is 10 m, before layer stacking. The bands that are only included in the layer stacking are Bands 2, 3, 4, and 8. In addition, NDVI and NDWI were generated. The IfSAR DTM, which has a resolution of 5 m, was resampled to 10 m to match the remainder of the bands. After that, all the bands with a resolution of 10m are layer stacked to generate a single image for classification.

3.4 Image Classification Object-Based Image Analysis (OBIA)

Object-based image analysis (OBIA) is a sub-discipline of geographic information science (GIScience) that focuses on partitioning remote sensing (RS) imagery into meaningful picture objects and analyzing their features on a geographical, spectral, and temporal scale (Hay and Castilla, 2006). OBIA requires image segmentation, classification, and ability to link individual objects. In OBIA, segmentation is the initial step, which involves partitioning an array of pixels into objects of a specific size based on the homogeneity of the pixels in the layers or bands under consideration. The segmentation is done in eCognition, a well-known OBIA software, where segmentation parameters such as item size, compactness, and shape are more defined and can be changed according to preferences.

Support Vector Machine (SVM) is one of the techniques that may be used to categorize objects in an image. It can handle multidimensional features and has been shown to produce accurate classification. The SVM training technique finds a hyperplane that divides the dataset into a discrete number of classes that is consistent with the training examples (Vapnik, 1979). The classification is done using a linear SVM, which assumes that multispectral land cover features can be separated linearly in feature space (Cortes & Vapnik, 1995). One might use numerous parameters in the classification using the eCognition program, such as the standard deviation of the layer values and the textural properties (homogeneity and entropy) of the objects, in addition to the mean layer values. Google Earth photos and visual interpretation were used to accumulate training points. To assign objects as samples for the SVM training, the points are imported to the map.

The classified base images are then merged and refined through contextual editing after the base image/s have been classified. Because of the image resolution, some features, such as built ups, bare land, and crop lands, are still misclassified due to their similar reflectance and properties. Also, some forest and fruit-bearing trees, as well as shrubs and trees. The resulting resource map and its accuracy are improved by correcting these misclassifications.

4. RESULTS AND DISCUSSIONS

4.1 Accuracy Assessment

Accuracy assessment is a significant part of any classification project. It compares the classified image to another data source that is accurate or ground truth data. Ground truth can be collected in the field; however, this is time consuming and expensive. The most common way to assess the accuracy of a classified map is to generate a set of random points from the ground truth data and differentiate that to the classified data in a confusion matrix.

The 2021 land cover was derived from land satellite images available in Google Earth Engine. Since this was processed remotely, it needs to be validated to check the input data’s integrity. Accuracy assessment is an important part of any classification project by comparing the classified image to another data source that is ground truth or considered to be accurate. Ground truth data can be collected in the field, which is time consuming and expensive, but ground truth data can also be derived from interpreting a high-resolution satellite imagery, existing classified imagery, or GIS data layers. ArcMap was used to perform the most common way to assess the accuracy of a classified map. By creating a set of random points from ground truth data and comparing that to the classified data in a confusion matrix. One thousand five hundred forty three (1,543) random ground truth points were used for the 2021 Land Cover using the create accuracy assessment points tool. The random points as seen in Figure 5 were manually identified again to populate the ground truth field.

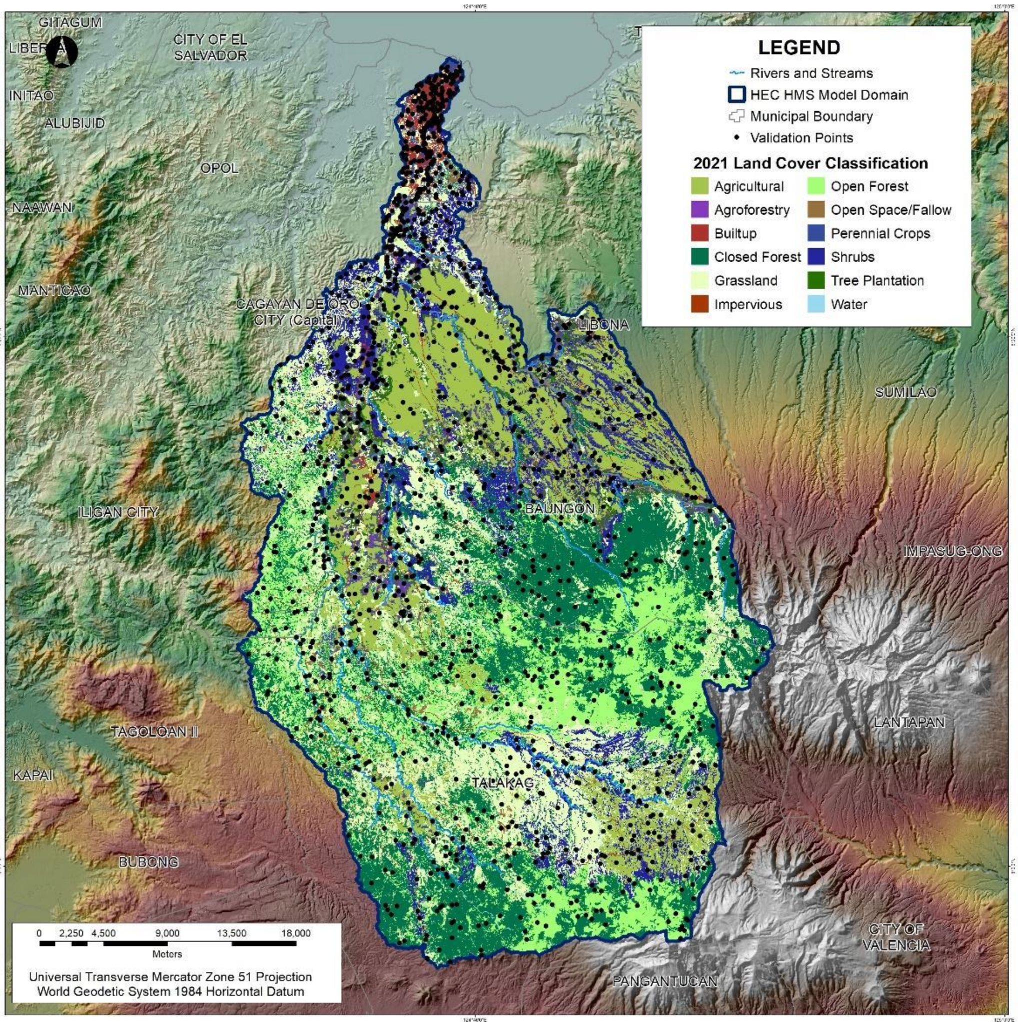

Figure 5. Validation Points for Accuracy Assessment of the 2021 Land Cover.

To compare and compute the accuracy of the classification, we used the Compute Confusion Matrix Tool. Those two tools ensure that each point has valid class values for the classified and ground truth fields. The tool computes a confusion matrix with errors of omission and commission and derives a kappa index of agreement and an overall accuracy between the classified map and the reference data. The 2021 Land Cover Classification garnered a user's accuracy of 95.01% and Kappa index of 94.31%. User's accuracy shows false positives, where pixels are incorrectly classified as a known class when they should have been classified as something else and also known as errors of commission, the kappa index of agreement gives an overall assessment of the accuracy of the classification.

4.2 2021 Land Cover Classification

The validated land cover was reclassified into twelve (12) different land cover classification. Each class's area is calculated in square kilometers as shown in Table 1.

Table 1. 2021 Land Cover Classification and Area

Land Cover Class	Area (in sq.km.)
Agricultural	225.072197
Agroforestry	21.698483
Built-up	19.791949
Closed Forest	327.827074
Grassland	261.881756
Impervious	14.467847

Open Forest	258.880841
Open Space/Fallow	19.61707
Perennial Crops	27.640607
Shrubs	124.469434
Tree Plantation	33.446238
Water	20.182706

Figure 6 represents the proportioning of the area distribution of the land cover class. The largest area is covered by the Closed Forest class and followed by the Open Forest. By default, the 2021 land cover is dominated by forest classification. Closed forest class can be visualized in the middle and in the upstream most part of the basin as reflected in Figure 7. While the Open Forest is scattered in the middle and few in the upstream part of the basin. Interestingly, grassland and agricultural ranked third and fourth as reflected in Table 1.

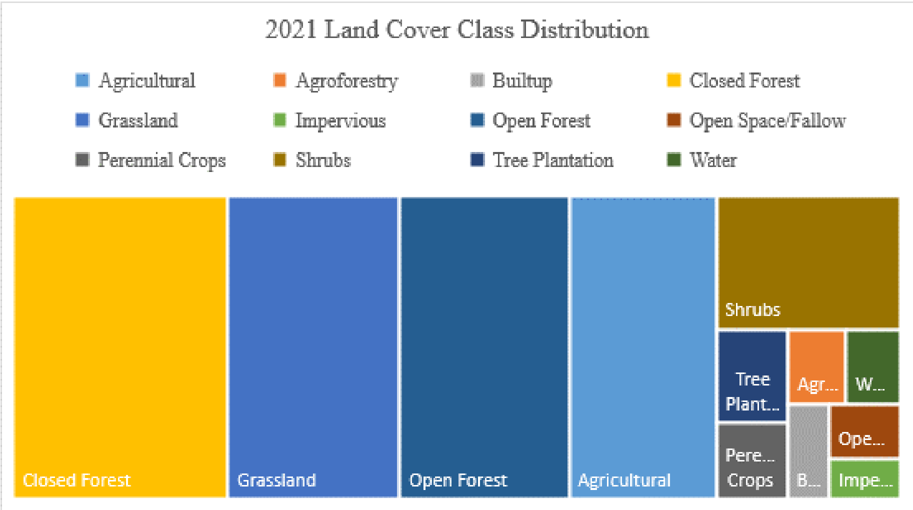

Figure 6. 2021 Land Cover Area Distribution

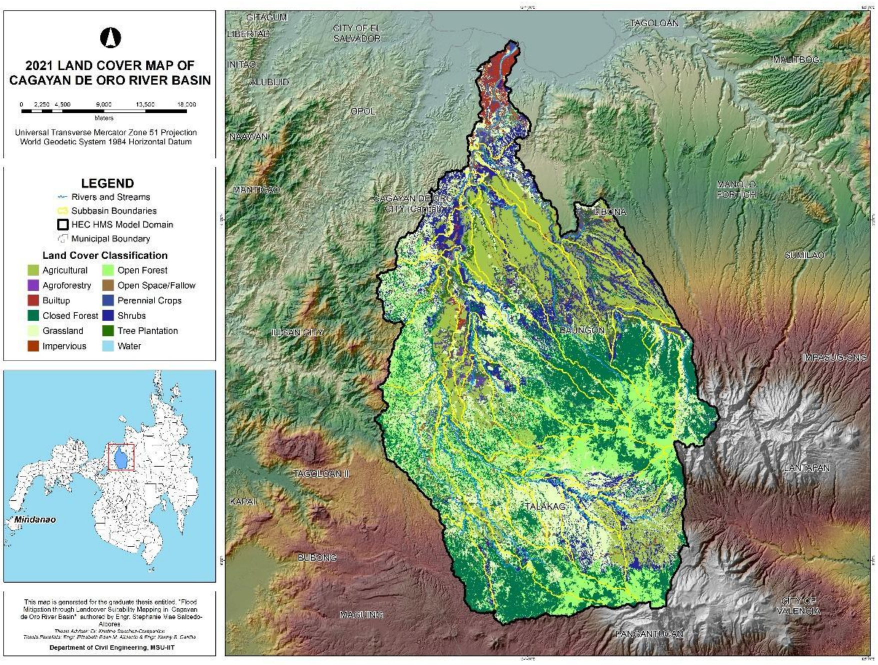

Figure 7. 2021 Land Cover Map of Cagayan de Oro River Basin

5. CONCLUSION

In summary, the findings and outcomes of this study underscore the remarkable effectiveness of Google Earth Engine as a valuable tool for the extraction and mapping of land cover within the Cagayan de Oro River Basin. The observed high accuracy in the land cover classification process not only validates the robustness of this platform but also highlights its immense potential for large-scale land cover analyses. The significance of these results extends beyond the boundaries of this specific study area. The demonstrated capability of Google Earth Engine holds immense promise for broader applications in environmental monitoring and resource management. Its capacity to process extensive datasets, analyze complex land cover patterns, and deliver precise classification outcomes positions it as a pivotal asset for decision-makers, researchers, and environmental agencies engaged in addressing the multifaceted challenges of land use and land cover changes.

By leveraging Google Earth Engine's capabilities, stakeholders can gain invaluable insights into the dynamic and evolving nature of land cover within the Cagayan de Oro River Basin and similar regions worldwide. This, in turn, empowers them to make informed decisions related to land management, conservation efforts, urban planning, and natural resource allocation. In conclusion, this study illuminates the transformative potential of Google Earth Engine in advancing our understanding of land cover dynamics, ultimately contributing to more sustainable and informed approaches to environmental stewardship and resource utilization.

ACKNOWLEDGEMENTS

This thesis study is anchored on the extension project of MSU-IIT with the title, “Development of a Decision Support Tool (DST) for the Cagayan de Oro River Basin”. The project is funded by Wetlands International, VEI (Dutch Water Operators), and Hineleban Foundation. This is conducted as a part of the feasibility study prior to the launching of a rehabilitation project by the Cagayan de Oro River Basin Management Council in partnership with the different funding agencies. With this, the author would like to thank Peter D. Suson, DiSDS, Engr. Elizabeth Edan M. Albiento and the MSU-IIT team of researchers for the opportunity and support they have provided for the conduct of the study. Most especially, to Dr. Kristine S. Companion, for taking this topic under her advisory and providing a fresh perspective on the objectives and processes in this study.

REFERENCES

Aplin, P. (2004). Remote sensing: land cover. Progress in Physical Geography: Earth and Environment, 28(2), 283–293. https://doi.org/10.1191/0309133304pp413pr

Cortes, C, & Vapnik, V (1995). Support-vector networks. Machine learning, Springer, https://doi.org/10.1007/BF00994018

Ghorbanian, A, Kakooei, M, Amani, M, Mahdavi, S, & ... (2020). Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated …. ISPRS Journal of …, Elsevier, https://www.sciencedirect.com/science/article/pii/S0924271620302008

Google. (2019.). Google Earth Engine. Retrieved from Google: https://earthengine.google.com/

Hay, GJ, & Castilla, G (2006). Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). Proc. 1st Int. Conf. OBIA, Citeseer, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=00f56c5d2ca1389921df92d6b284f4d663d5dc25>

Petchprayoon, P, Blanken, PD, & ... (2016). A decade of changing surface energy balance components over a large water region. Remote Sensing of …, spiedigitallibrary.org, <https://doi.org/10.1117/12.2241516.short>

Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *MPDI.*

Sahebjalal, E., & Dashtekian, K. (2013). Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods. African Journal of Agricultural Research, 4614-4622.

Santillan, J. R., Amora, A. M., Makinano-Santillan, M., Gingo, A. L., & Marqueso, J. T. (2019). Analyzing the Impact of Land Cover Change to the Hydrologic and Hydraulic Behaviours of the Philippines' Third Largest River Basin. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4.

Vapnik, VN (1979). Introduction to statistical learning theory. IEEE transactions on neural networks