LAND COVER CHANGE ANALYSIS OF CARRASCAL, MADRID AND CANTILAN MINING SITES IN SURIGAO DEL SUR USING REMOTE SENSING

G. A. M. Narciso [1]*, J. A. Principe [2]

[1] National Graduate School of Engineering, College of Engineering, University of the Philippines Diliman - narciso.gilson@gmail.com

[2] Department of Geodetic Engineering, College of Engineering, University of the Philippines Diliman - japrincipe@up.edu.ph

KEY WORDS: Change Detection, Land Cover, Environmental Monitoring, Image Classification, Multitemporal

ABSTRACT:

The Philippines is recognized as amongst the natural resource-rich countries which has been evident in the influx of extractive industry investors, most specifically in the mining industry. An example of this is the case of mining sites in the municipalities of Carrascal, Madrid, and Cantilan (CMC) in Surigao del Sur region of the Philippines. In this region, development of mining areas was significantly noticeable in a span of seven (7) years since 2015. With this being an alarming case due to its possible negative impacts on the surrounding communities, this study developed an open-source monitoring system of mining activities using remote sensing technologies employing supervised image classification using the Random Forest algorithm. This study is mainly motivated by the need for a monitoring system to ensure responsible mining practices in the country, most especially since there has been an increasing global clamour for a more environmental-friendly and climate-sensitive and sustainable practices. The methodology developed for this study utilized a post-classification change detection approach in which multitemporal land cover (LC) models were produced to capture the land cover changes. To model the LC changes, land cover models for 2015 and 2022 were produced obtaining overall accuracies between 0.93 to 0.95, respectively. Using these models, changes due to mining development were delineated through a multitemporal overlay analysis approach. This change detection model obtained an F1-score of 0.93 indicating high accuracy and satisfactory performance of the model and consequently, the its potential as a mining site monitoring system to achieve the goal of ensuring responsible mining industry practice in the country.

1. INTRODUCTION

The Philippines is one of the world's major sources of gold, chromite, copper and nickel (Domingo, E. G. 1993; MICC, 2022). In the 1970s, the mining industry has contributed to as much as 1.4 percent of the country's gross domestic product (GDP) (MICC Policy Note, 2022). Although, this is considerably minimal, at the regional level, the mining industry allowed some regions in the country to experience significant economic growth (MICC Policy Note, 2022). However, since the 1980s, a decline in the industry has been brought about by the country's environmental sensitivity and awareness. A major government intervention was the suspension of 75 Mineral Production Sharing Agreements (MPSA) to protect the country's major watersheds (MICC, 2022; Simeon, 2021; DENR). The potential of the mining industry in the country is yet to be realized however, its impacts, most especially surface mining, on the people and the environment has to be addressed. This therefore necessitated an objective and science-based evaluation of mining operations (MICC, 2022; Madasa, 2021).

1.1 Goals and Objectives

To address the issues of mining industry, this study proposes the use of geospatial technologies, particularly remote sensing, to develop a monitoring tool. Remote sensing offers an efficient, scientific and inexpensive alternative to gain multi-temporal information about the Earth's surface using satellite data.

The objectives of this study are as follows:

- Develop land cover models using machine learning to characterize the study area;
- Explore the use of remote sensing derived geospatial indices;
- Utilize post-classification change detection to identify land over change due to mining activities

1.2 Significance

In 2016, the Mining Industry Coordinating Council (MICC) was prompted to implement a more thorough evaluation of existing mine operations in the country after revealing from the DENR mine audit the violations of 30 out of 41 existing mines (MICC, 2022). This led to supporting an objective and scientific approach to evaluate mining operations (MICC, 2022). Therefore, the development of a monitoring tool for mining activities using remote sensing could offer an efficient, scientific, timely and cost-effective approach to evaluate mining operations by gaining insights about the surface dynamics as detected by satellites. Moreover, various satellite data derivatives could help quantitatively assess the mining operations.

1.3 Related Studies

This project builds on the works of Madasa et al. (2021) whoh developed a remote sensing framework to assess land use and land cover changes brought about by the development of mining sites in South Africa. Their study however was focused in assessing the application of different geospatial indices in

* Corresponding author

improving a land cover image classification model using Maximum Likelihood classifier (Madasa et al., 2021). With regards the geospatial indices, Normalized Difference Indices for vegetation (NDVI), water (NDWI), built-up (NDBI), soil (NDSI) and the Global Environmental Monitoring Index (GEMI) were assessed in their study to help in discriminating general land cover types such as vegetation, water, built-up and soil features which resulted in land cover models accuracies from 88% to 96% (Madasa et al., 2021). Although the study of Madasa et al. produced high accuracy values, this study aims to develop it further by utilizing current machine learning techniques.

The study of Fargas et al. (2021) which developed an agri-urban land conversion monitoring using post classification change detection technique was also adopted in this study. Their study utilized Random Forest algorithm to produce LULC maps of their agriculture-dominant area for 3 different years and agri-urban conversion was then segmented through a boolean filtering approach between the pre- and post- LULC models (Fargas et al., 2021). The boolean filtering process identified pixels classified as agricultural from the pre LULC model which were then compared to their corresponding class in the post LULC model (Fargas et al., 2021). Pixels which transitioned from agriculture to either built-up or bare soil were segmented and identified as agri-urban conversion (Fargas et al., 2021).

With regards the use of machine learning for land use and land cover mapping applications, machine learning algorithms such as Random Forest and Support Vector Machines (SVM) are generally utilized. In the study of Talukdar et al. (2020), machine learning algorithms for image classification were tested which resulted in Random Forest (RF) obtaining highest accuracy of 0.89 which suggests that it is the best LULC classifier. Moreover, aside from being non-parametric and flexible to highly dimensional data, RF could produce consistent accuracies compared to other algorithms (Talukdar et al., 2020; Torbick et al., 2016; Pelletier et al., 2016; Fargas et al., 2021). However, in the study of Basheer et al. (2022), SVM was reported to perform better than RF in terms of overall accuracy.

2. METHODOLOGY

2.1 Study Area

The CMC Mining Sites in Surigao del Sur experienced significant expansion of the existing mining tenements in the region. This area is covered by the municipalities of Carrascal, Madrid and Cantilan in the province of Surigao del sur. In these municipalities, Mineral Production Sharing Agreements (MPSA) were granted to three nickel mining corporations, the Carrascal Nickel Corporation, Marcventure Mining and Development Corporation (MMDC) and the Madrid Surigao del sur Nickel Mining Corporation.

The case of CMC Mining Sites is found to be a suitable case for this study due to the environmental issues which can be found in the region such as the overlapping of MPSA tenements with a National Integrated Protected Area Systems (NIPAS) in Surigao, as well as the adjacency of the mining sites to agricultural lands and to resource-rich coastal areas.

Figure 1. CMC overlaid with existing mining tenements obtained from MGB's mining tenements control map and NIPAS obtained from NAMRIA's Geoportal.

Moreover, inspecting the study area in Google Earth Pro, it can be seen that there was significant change in the area due to mining expansion in the last seven years which should be subjected to active monitoring given its possible implications to its surrounding ecosystems.

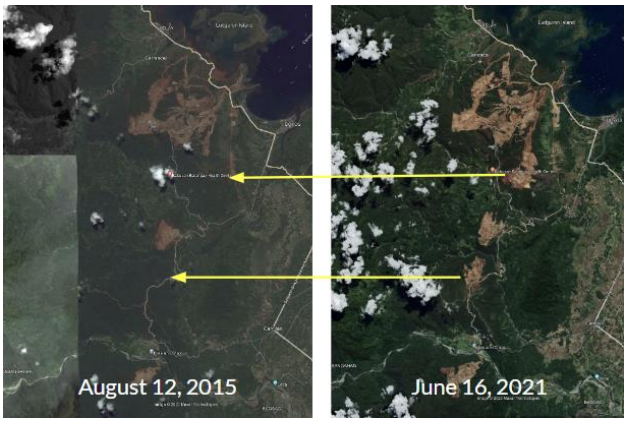

Figure 2. High resolution satellite images of the study area for 2015 and 2021.

2.2 Data

This study used 2015 and 2022 Landsat 8 images covering the study area. Annual cloud-free images were produced by masking out cloud-covers from the images and mosaicking all images by computing the median value at pixel level.

Aside from optical images, the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) was also used to derive elevation and slope data to account for the study area's topography.

Table 1 shows a summary of indices used in this study which were mainly adopted from the study of Madasa et al., (2021).

Index	Equation
NDVI	(NIR – Red) / (NIR + Red)
NDWI	(Green – NIR) / (Green + NIR)
NDBI	(SWIR – NIR) / (SWIR + NIR)
BSI	((Red + SWIR) - (NIR + Blue)) / ((Red + SWIR) + (NIR + Blue))

GEMI	Xi = ((NIR^2 - Red^2) * 2 + (NIR * 1.5) + (Red * 0.5)) / (NIR + Red + 0.5). GEMI = Xi * (1 - (Xi * 0.25)) - ((Red - 0.125) / (1 - Red))

Table 1. Summary of equations of geospatial remote sensing indices.

2.3 Image Classification

Random Forest (RF) algorithm was used due to its flexible nature allowing it to handle non-normalized and high dimensional data (Fargas et al., 2021; Pelletier et al., 2016). Moreover, the algorithm can provide insight about variable importance which would help in better assessing the different geospatial indices used in this study.

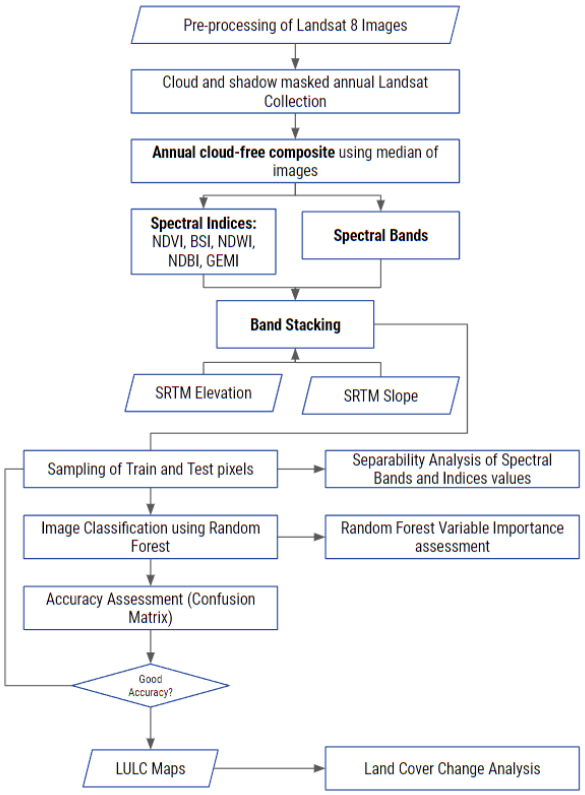

Figure 3. Land cover modelling methodology using Random Forest classifier.

Figure 3 shows a flow chart of the methodology for this study which simply follows a straightforward image classification process using Random Forest classifier. A significant feature of this method, however, is the stacking of spectral bands, geospatial indices, elevation and slope data into one variable which aims to improve the classification framework of Madasa et al. (<year>) by creating a variable which are supposed to capture the differences of the different land cover features.

With regards the training of a classifier, train and test pixels were selected using both the before and after mining development Landsat images in Google Earth pro. 'Unchanged' areas were mainly sampled to ensure that spectral responses were consistent. 200 pixels per class were sampled which were then divided into train and test set. 70% and 30% distribution were used for train and test, respectively.

Although in most studies it is suggested to pool a large number of train pixels to ensure model accuracy, to optimize the workflow, the study of Foody and Mathur in 2004 was followed in which they argued that it is not necessary to select a huge training size. In fact, their study found out that machine learning algorithms, such as SVM, are capable of learning from a limited number of training pixels. In their study, out of 150 train pixels, only 37 pixels were used to delineate hyperplanes which act as boundaries to differentiate one class from another (Foody and Mathur, 2004).

3. RESULTS AND DISCUSSION

One of the objectives of this study is to implement the method using an open-source platform to ensure the accessibility of the developed framework for possible future use. In this case, Google Earth Engine (GEE) was used to access the data and to implement the methodology of this study. GEE is an open-source platform which has access to repositories of different open-source satellite data such as Landsat, Sentinel and MODIS (Gorelick et al., 2017). It also has built-in remote sensing functions which can be implemented to automate remote sensing processes (Gorelick et al., 2017).

3.1 Cloud-free Composite Images

For this study, annual cloud-free composite for years 2015 and 2022 were produced using GEE. The following figure shows the produced cloud-free images. These images were produced by collecting a year worth of Landsat 8 images covering the study area. The image collection was then subjected to cloud-masking using Landsat's QA pixels to segment cloud covers. After which, the image collection was aggregated into one image using the median function which selects the median cloud-free value at pixel level producing a cloud-free composite image of the study area.

Figure 4. Annual cloud-free composite of CMC region using Landsat 8 Image collections for 2015 and 2022.

3.2 Land Use Land Cover Models

Figures 5 and 6 show the LULC models for years 2015 and 2022 respectively. The selected land cover classes for this study include cropland, bare soil, built-up, forest, grassland and water features. Croplands in this case include cultivated and uncultivated areas. Grasslands pertains to both grasslands and shrublands. For bare soil, it includes open barren areas and mining sites.

The following tables summarize the overall, kappa, producer and users' accuracies for each model which were computed from the confusion matrix derived in GEE.

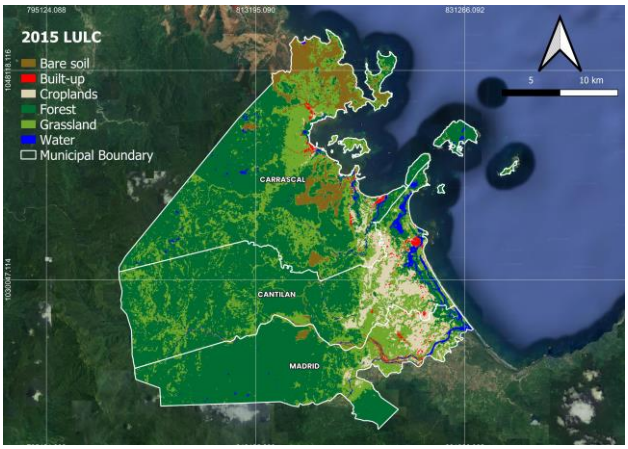

Figure 5. The 2015 land cover distribution in the CMC Region.

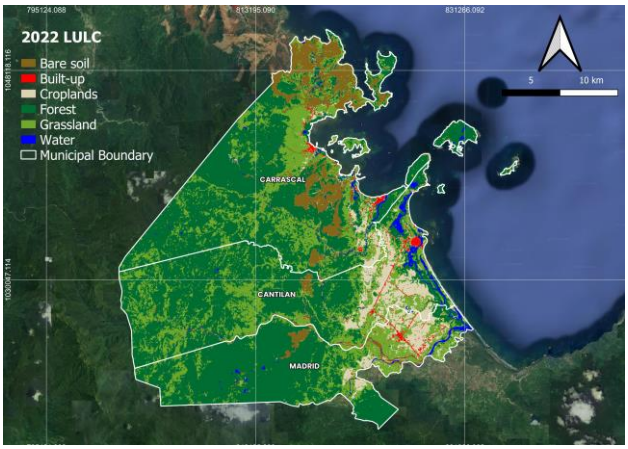

Figure 6. The 2022 land cover distribution in the CMC Region.

Based on the test data, the models have significantly high overall accuracies and kappa statistics ranging from 0.93 to 0.95. The producer and user accuracies also obtained high values. Among the classes, the model had most difficulty classifying built-up features obtaining lowest producer and user accuracies of 0.86 and 0.88 respectively for both years. Although the values are still sufficiently high. This implies that the trained models were able to sufficiently differentiate the land cover classes set for this study.

Model	Overall Accuracy	Kappa Statistics
2015	0.94	0.93
2022	0.95	0.93

Table 2. Overall accuracy and kappa statistics of the land cover models.

Producer's Accuracy		
	2015	2022
Cropland	0.98	0.98
Bare Soil	0.92	0.95
Built-up	0.95	0.86
Forest	0.93	0.93
Grassland	0.89	0.98
Water	0.98	0.98

Table 3. Producer's accuracy of the land cover models.

User's Accuracy		
	2015	2022
Cropland	0.97	0.98
Bare Soil	0.97	0.97
Built-up	0.88	0.93
Forest	0.90	0.95
Grassland	0.95	0.90
Water	0.98	0.95

Table 4. User's accuracy of the land cover models.

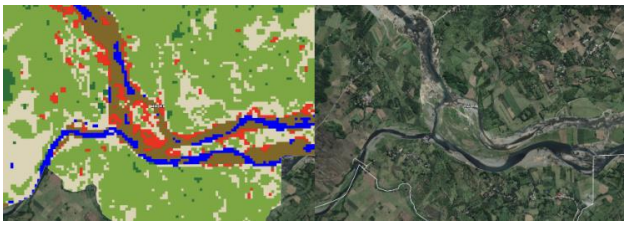

Figure 7. Bare-soil classified as built-up.

Figure 8. Cropland classified as grassland.

However, while the results of the confusion matrix for the 2015 and 2022 models suggest that they are considerably accurate, there were still misclassifications in the classified images which can be noticed through visual inspection. The previous figures show some examples of misclassifications especially for cases between bare soil and built-up and between cropland and grasslands.

Confusion in classifying these classes is mainly explained by their high similarity in spectral response and visual characteristics. Built-up and bare soil, specifically gravel-dominant areas have almost identical responses. Croplands on the other hand, depending on the season may look similar to either grasslands or bare soil features. Nevertheless, land cover classification models developed using Random Forest produced considerably high results.

3.3 Separability Analysis

For this section, separability of the different land cover features is assessed based on the train pixels collected spectral and geospatial index information. The following graphs are collection of the box plots for each land cover class for each variable used in the LULC modelling.

Using spectral bands only, water and forest features are easily separable. For the other classes, significant overlap was found between built-up and bare soil and between cropland and grasslands. Among these classes, built-up class has the widest spectral response range in the RGB bands due to the wide range of colour characteristics of built-up features.

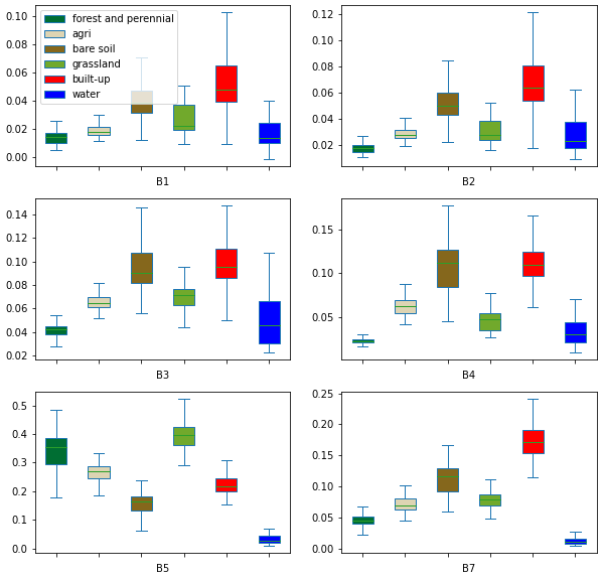

Figure 9. Comparison of responses per class per spectral band using the train data summarized using boxplots.

The significant overlap between most of the land cover classes highly suggested the use of derived variables such as geospatial indices. With that, five indices were explored in this study, NDVI, NDWI, BSI, NDBI and the GEMI.

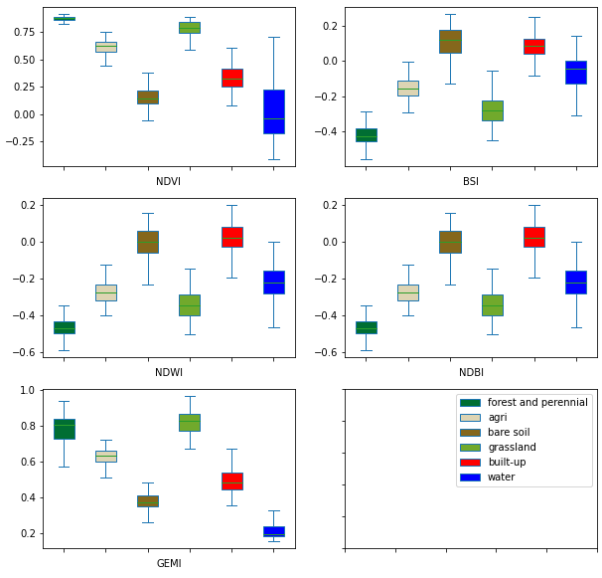

Figure 10. Comparison of geospatial index value per class using the train data summarized using boxplots.

Among these indices, GEMI and NDVI were able to discriminate the four confusing classes. In terms of separating croplands from grasslands, GEMI exhibited best performance. NDVI on the other hand have shown similar capabilities in terms of discriminating bare soil and built-up features.

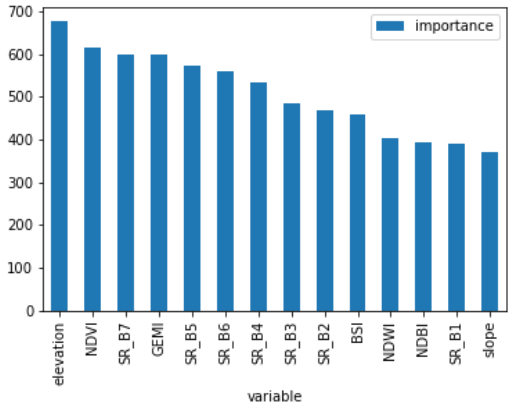

Figure 11. Variable importance graphs for the 2015 (top) and 2022 (bottom) land cover models.

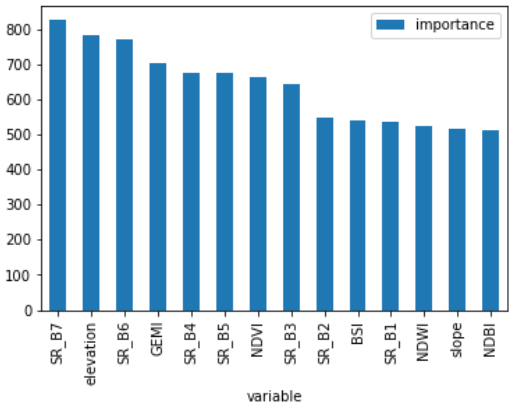

Figure 12. Variable importance graphs for the 2015 (top) and 2022 (bottom) land cover models.

Consistent to the findings of separability analysis are the contributions of GEMI and NDVI to the Random Forest image classification models. Figures 11 and 12 show that for both the 2015 and 2022 models, GEMI and NDVI performed better compared to the other indices. NDBI, NDWI and BSI were ranked among the lowest implying their low contribution to the models. GEMI and NDVI are among the top 10 variables which had significant contributions to the image classification process.

3.4 Change Detection Analysis

After generating the LULC models of the study area for 2015 and 2022, a change detection map was produced to indicate changes due to mining expansion. The red regions shown in the following map indicate converted vegetated lands to mining lands.

To delineate changes which were brought about by mining expansion, changes from vegetation to bare soil was set as the phenomena's main indicator for this study. By identifying areas which were converted from vegetation to bare soil, the researchers were able to produce a change detection map that is supposed to model the changes brought about by the development of the mining sites in the region.

This result was assessed using a confusion matrix as well. 100 pixels were sampled for changed areas and another 100 pixels were sampled for unchanged areas which were used to detect true and false positives and true and false negatives.

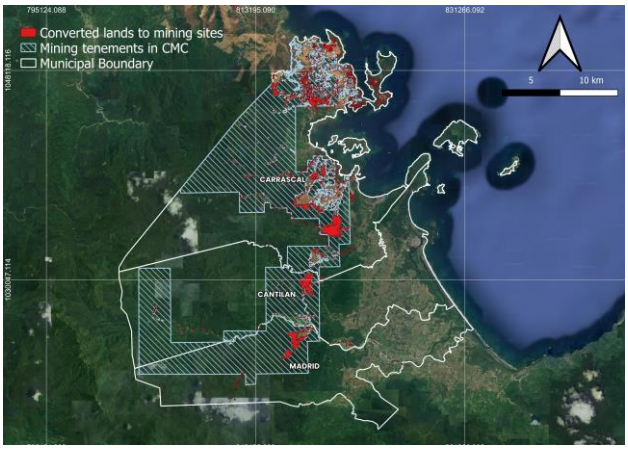

Figure 13. Change detection map of Carrascal, Madrid and Cantilan from 2015 to 2022.

The following confusion matrix summarized the accuracy of the detected changes produced from the multi-temporal LULC models. For this study, the unchanged region was limited to the mining tenement regions only as this implies likeliness of experiencing land cover change.

Based on the confusion matrix, the change detection map was able to sufficiently identify land cover changes brought about by the development of mining sites in the area with an overall accuracy of 0.94. User and producer accuracies ranged from 0.88 to 0.99 indicating high capability to detect land cover changes.

		Test Data		
Classified		Changed	No Change	User's Accuracy
	Changed	88	12	0.88
	No Change	1	99	0.99
	Producer's Accuracy	0.99	0.89	0.94

Table 5. Change detection accuracy matrix.

The F1-score for change detection was also computed using the model's precision and recall which mainly pertain to the Changed class user and producer accuracies respectively. The model obtained an F1-score of 0.93

3.5 Land Cover Change Analysis

Class	2015		2022		Diff.
	Area (ha)	%	Area (ha)	%	
Croplands	4505.19	7.64	5,430.82	9.21	1.57
Forest	37444.00	63.50	35,500.00	60.26	-3.25
Water	992.39	1.68	872.63	1.48	-0.20
Built-up	552.80	0.94	500.40	0.85	-0.09
Grassland	12530.98	21.25	12,700.00	21.53	0.28
Bare soil	2938.84	4.98	3,935.87	6.68	1.69

Table 6. Land cover distribution values for 2015 and 2022.

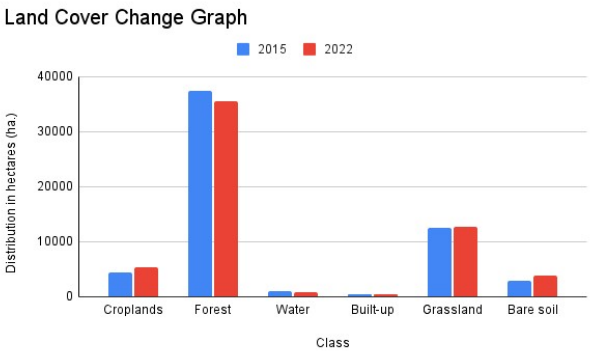

Figure 14. Land cover distribution for 2015 and 2022 in CMC region.

The land cover models of 2015 and 2022 captured the increase in bare soil which is mainly due to the development of mining in the area. Within 7 years, almost 1,073 hectares of vegetated areas were mined and converted to bare soil which indicated mining expansion.

Aside from this, a noticeable change was the decrease of forest distribution which may be explained by the increase in classification of forest as grasslands.

The increase and decrease in distribution of croplands and built-up, respectively can be attributed to the misclassification between croplands and grasslands and between bare soil and built-up features.

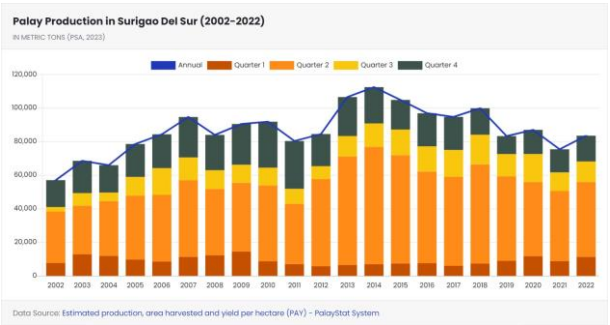

Figure 15. Palay production trend graph of Surigao Del Sur from 2002 to 2022 obtained from PhilRice Palay Stat website (https://palaystat.philrice.gov.ph/).

Looking at the possible impacts of this change, rice production data for Surigao Del Sur from 2002-2022 was obtained from the Palay Stat website of PhilRice. From 2015 to 2022, a decrease in rice production was observed. This may be attributed to the significant increase in mining activity in the province although it would require further studies since the trend in croplands indicates otherwise.

From 2015 to 2022, there is an observed increase in the distribution of croplands however, this may not accurately represent the actual use of land since this study uses one image for a year only and croplands may exhibit variability within a year. Using a single image only to model such land use is indicative and is susceptible to misclassification of croplands as either bare soil or grassland, and vice versa which could explain the distribution trend between 2015 and 2022.

4. CONCLUSION

This study was able to develop a monitoring system using open-source technology for change detection about brought by mining development. For the satellite imagery, Landsat 8 images were utilized for land cover modelling and representation. Google Earth Engine, an open-source remote sensing platform, was also leveraged in this study proving the possibility of an accessible monitoring system which can be used by the authorities (Gorelick et al., 2017).

With regards the land cover change modelling, multi-temporal land cover maps were generated using Random Forest approach. Although the generated maps were susceptible to misclassifications especially between bare soil and built-up and between croplands and grassland, which are mainly attributed to the high spectral response similarity between these classes, the multi-temporal models were still able to capture the changes due to mining development in the region.

Moreover, indices such as NDVI, NDWI, NDBI, BSI and GEMI were included in the generation of land cover maps. Among these geospatial indices, NDVI and GEMI have shown good performance in terms of contribution in the land cover modelling. In discriminating ambiguous features, both indices were also able to separate croplands from grasslands and built-up from bare soil.

Overall, the study showed potential in terms of developing a monitoring system for mining sites. Land cover change analysis can therefore be implemented using machine learning and open-source remote sensing data.

RECOMMENDATIONS

The main challenge encountered in this study was discriminating ambiguous features such as croplands, grasslands, bare soil and built-up. For croplands, generally, the use of land for agricultural purposes can be challenging to capture most especially when using a single image analysis such as the case of this study. Integration of a multi-temporal dataset to capture phenological properties of land cover classes can be implemented to further improve discrimination of such land cover class. The use of SAR data, which is cloud penetrating, therefore not susceptible to cloud covers, can be modelled in such a way that it is capturing phenological characteristics of surface features.

With regards to remotely sensed data, aside from Landsat 8, Sentinel 2 MSI is another open-source satellite imagery which has higher spatial and temporal resolution. Creating cloud-free composites may be better with the use of Sentinel 2 data for more recent analysis.

Lastly, since a multi-temporal analysis is being implemented in this study, relative calibration of images could also be implemented such as histogram matching. This way, the monitoring system can be made more robust and adaptive to localize implementations.

REFERENCES

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., & Liu, S. (2022). Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques. Remote Sensing, 14(19), 4978. https://doi.org/10.3390/rs14194978

Department of Environment and Natural Resources. nd. 75 MINING PERMITS FACE CANCELLATION. Retrieved from: https://ncr.denr.gov.ph/index.php/news-events/press-releases/75-mining-permits-face-cancellation

Domingo, E. G. (1993). The Philippine mining industry: status and trends in mineral resources development. Journal of Southeast Asian Earth Sciences, 8(1–4), 25–36. https://doi.org/10.1016/0743-9547(93)90004-9

Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sensing of Environment, 93(1–2), 107–117. https://doi.org/10.1016/j.rse.2004.06.017

Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sensing of Environment, 103(2), 179–189. https://doi.org/10.1016/j.rse.2006.04.001

Fargas., D.C., Narciso, G. A. M., & Blanco, A. C. (2021). MONITORING AND ASSESSMENT OF AGRI-URBAN LAND CONVERSION USING MULTI-SENSOR REMOTE SENSING AND GIS TECHNIQUES. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3–2021, 117–124. https://doi.org/10.5194/isprs-annals-V-3-2021-117-2021

Liu, W., Qin, R., & Su, F. (2019). Weakly supervised classification of time-series of very high resolution remote sensing images by transfer learning. Remote Sensing Letters, 10(7), 689–698. https://doi.org/10.1080/2150704X.2019.1597295

Madasa, A., Orimoloye, I. R., & Ololade, O. O. (2021). Application of geospatial indices for mapping land cover/use change detection in a mining area. Journal of African Earth Sciences, 175, 104108. https://doi.org/10.1016/j.jafrearsci.2021.104108

Mining Industry Coordinating Council. 2022, April. Review of Philippine Large-Scale Metallic Mines: Going Beyond Compliance Towards Sustainability. Retrieved from: https://www.dole.gov.ph/php_assets/uploads/2022/06/MICC-Mining-Policy-Note-Online-Version.pdf

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment, 187, 156–168. https://doi.org/10.1016/j.rse.2016.10.010

Simeon, L.M. 2017. DENR cancels 75 MPSAs in watersheds. Retrieved from: https://www.philstar.com/headlines/2017/02/14/1672012/denr-cancels-75-mpsas-watersheds

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-Use Land-Cover

Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing, 12(7), 1135. https://doi.org/10.3390/rs12071135

Torbick, N., Chowdhury, D., Salas, W., & Qi, J. (2017). Monitoring Rice Agriculture across Myanmar Using Time Series Sentinel-1 Assisted by Landsat-8 and PALSAR-2. Remote Sensing, 9(2), 119. https://doi.org/10.3390/rs9020119