DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

IJES
Indonesian Journal of Environmental Sustainability

Published by :
Center For Environmental Studies
UIN Ar-Raniry Banda Aceh

EXPLORING THE ROLE OF SPECTRAL INDICES ON IMPROVING LAND COVER CLASSIFICATION ACCURACY BASED ON SENTINEL-2 SATELLITE IMAGERY IN BANDA ACEH CITY, INDONESIA

Mulkal[1*], Putra Rizkya[2], Nurul Afla[1], Muchlis[2]

[1]*Department of Mining Engineering, Syiah Kuala University, Banda Aceh, Indonesia*
[2]*Department of Regional and City Planning, Syiah Kuala University, Banda Aceh, Indonesia*

***Email Correspondence:** mulkal@usk.ac.id*

Received: 23 October 2024 Accepted: 23 December 2024 Published: 31 December 2024

ABSTRACT

The land cover on the Earth's surface is constantly changing due to natural and human activities such as settlements, agriculture, mining, natural hazards, and more. These changes will continue as long as life exists on Earth, making land cover change monitoring a never-ending task. Land cover classification involves defining the existing land cover on the Earth's surface using satellite imagery data. Random Forest is a popular classification algorithm used in remote sensing. The aim of this research is to determine the role of spectral indices in improving land cover accuracy using the random forest method. Twelve spectral indices were used, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), Global Environment Monitoring Index (GEMI), Urban Index, Advanced Vegetation Index (AVI), Normalized Built-up Area Index (NBAI), Modified Bare Soil Index (MBI), and others. By combining these spectral indices with the 12 Bands of Sentinel-2 Satellite Imagery, the accuracy of land cover classification increased from 87% to 91%. The results showed that NBAI played a more important role compared to other spectral indices, with 6%, followed by NDVI with 5.6% and Urban Index with 4.9%.

Keywords: land cover, spectral index, random forest

ABSTRAK

Perubahan tutupan lahan akan terus terjadi seiring dengan aktivitas alam dan manusia seperti permukiman, pertanian, pertambangan, bencana alam dan lain sebagainya. Salah satu metode yang umum dilakukan dalam identifikasi tutupan lahan adalah metode penginderaan jauh dengan melakukan klasifikasi nilai piksel citra satelit. Salah satu algoritma yang populer digunakan dalam proses klasifikasi tutupan lahan adalah Random Forest. Tujuan dari penelitian ini adalah untuk mengetahui peran indeks spektral dalam meningkatkan akurasi tutupan lahan dengan menggunakan metode Random Forest dengan menggunakan citra satelit Sentinel-2. Dalam penelitian ini digunakan dua belas indeks spektral yaitu: Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), Global Environment Monitoring Index (GEMI), Indeks Perkotaan, Advanced Vegetation Index (AVI), Normalized Built-up Area Index (NBAI), Modified Bare Soil Index (MBI), dan lain-lain. Hasil penelitian ini menunjukkan penggunaan indeks-indeks spektral tersebut dengan 12 Band Citra Satelit Sentinel-2 dapat meningkatkan akurasi klasifikasi tutupan lahan dari 87% menjadi 91%. Hasil penelitian juga menunjukkan bahwa indeks spektral NBAI memainkan peran yang lebih penting dibandingkan dengan indeks spektral lainnya, yaitu sebesar 6%, diikuti oleh NDVI sebesar 5,6% dan Indeks Perkotaan sebesar 4,9%.

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

Kata kunci: Tutupan lahan, Spektral Indeks, Random Forest

Introduction

Human activities on the earth's surface, such as settlement, agriculture, mining, and hazard mitigation, will inevitably lead to changes in the earth's surface and its characteristics over time. However, these changes can have adverse effects on the ecosystem and, in turn, on humans themselves. For instance, converting a forest into an oil palm plantation can threaten the habitat of native species, degrade soil quality, and increase the risk of flooding in low-lying areas. To mitigate these impacts, it is crucial to monitor changes in land cover and take measures to maintain a healthy environment and preserve the quality of the ecosystem.

Remote sensing has been widely utilized as a powerful tool in various applications such as land cover mapping, environmental monitoring, and natural resource management (Madasa et al., 2021). One of the crucial components of remote sensing is the use of satellite imagery, which captures vast areas of the earth's surface in various wavelengths of the electromagnetic spectrum. The satellite's bands provide rich information that can be used to identify and classify various land cover features. Furthermore, the derivation of spectral indices from these bands has been widely recognized as a useful technique to enhance the object classification process. Spectral indices are derived from mathematical operations on the values of the satellite's bands and provide additional information about the physical and chemical properties of the objects on the earth's surface.

The Random Forest method is one of the most widely used machine learning algorithms for land cover classification (Gislason et al., 2006). It has gained popularity due to its high accuracy and ease of use. The method builds a decision tree from the training data, and the results of these trees are combined to form the final classification result.

With this background, the objective of this study is to investigate the role of spectral indices in enhancing the accuracy of land cover classification results using the Random Forest method. The study will evaluate the importance of various spectral indices, including Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Soil-Adjusted Vegetation Index (SAVI), among others. The findings of this study will provide valuable insights into the importance of spectral indices in the land cover classification process and contribute to the development of more accurate and effective remote sensing applications.

Methods

The study area for this research is Banda Aceh municipality which is located in the most Western Indonesia. Banda Aceh is the capital city of Aceh Province. Banda Aceh is a medium city with some land cover namely: water body, fish pond, built area, unbuilt area, bare land, un-irrigated paddy field and forest which are grouped into several land cover types, such as water body, vegetation, open land, built area and unbuilt area.

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

In this research, multiple spectral indices were employed to improve land cover classification accuracy. These indices include Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), Global Environment Monitoring Index (GEMI), Urban Index, Advanced Vegetation Index (AVI), Normalized Built-up Area Index (NBAI), Modified Bare Soil Index (MBI), Normalized Difference Pond Index (NDPonI), Sentinel-2 Water Index (S2WI), Modified Normalized Difference Water Index (MNDWI), and Enhanced Modified Bare Soil Index (EMBI). The formulas for each spectral index can be found in Table 1.

Table 1. Spectral Indices Formula

Spectral Indices	Equation	Reference
NDVI	(N-R)/(N+R)	Rouse et al., 1974
NDWI	(G-N)/(G+N)	McFeeters, 1996
NDBI	(S1-N)/(S1+N)	Zha et al., 2003
GEMI	n(1-0.25n)-((R-0.25)/(1-R)) n=(2(N^2-R^2)+1.5N+0.5R)/(N+R+0.5)	Pinty and Verstraete, 1992
Urban Index	(S2-N)/(S2+N)	Kawamura et al., 1996
AVI	$(N(1.0-R)(N-R))^{(1/3)}$	Rikimaru et al., 2002
NBAI	((S2-S1)/G)/((S2+S1)/G)	Waqar et al., 2012
MBI	((S1-S2-N)/(S1+S2+N))0.5	Nguyen et al.,2021
NDPonI	(S1-G)/(S1+G)	Lacaux et al., 2007
S2WI	(RE1-S2)/(RE1+S2)	Jiang et al., 2021
MNDWI	(G-S1)/(G+S1)	Xu, 2006
EMBI	(MBI-MNDWI-0.5)/(MBI+MNDWI+1.5)	Zhao and Zhu, 2022

Where:

N = Near Infra Red Band
RE1= Red Edge 1 Band
S1= Short Wave Infra Red 1 Band
S2=Short Wave Infra Red 2 Band
G= Green Band
R= Red Band
B= Blue Band

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

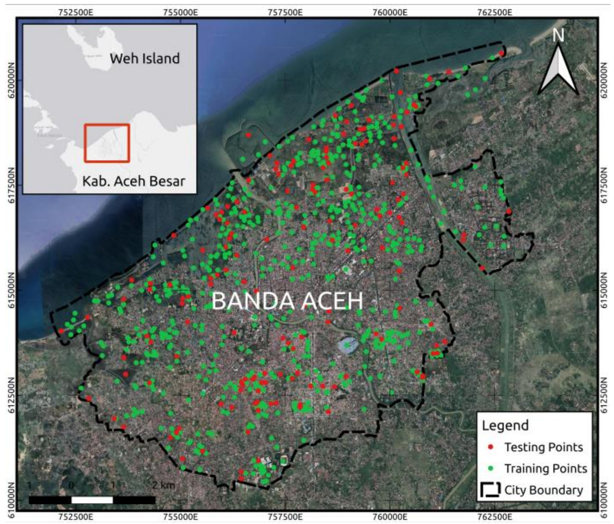

Figure 1. Training and Testing Points

Approximately 812 sampling points were manually selected based on High Resolution Satellite Imagery, with 650 points utilized for training and 162 points reserved for testing the dataset as shown in figure 1. The sample selection was performed through visual inspection and was based on the land cover descriptions found in Table 2, which comply with Indonesian National Standard 7645:2010 (BSNI, 2010).

Table 2. Land Cover Type Description

Land Cover Type	Description
Water Body	Part of earth surface that covers significant amount of water such as sea, lake, swamp, river, reservoir, fishpond and stream
Vegetation	Surface that has plants at least 4% within 2 months or covered with lichens/mosses more than 25% if no other vegetation type exist
Open land	Natural, semi-natural or artificial land which is not well covered by its characteristics and can be categorized as consolidated or unconsolidated surface.
Built Area	Area has been substituted with permanent or waterproof features like settlement, road, industrial zone, port, etc.
Unbuilt Area	Land with human intervention that eradicate original land cover type, although it’s not developed the same as built area.

The accuracy of classification results is measured with consumer accuracy, producer accuracy and overall accuracy. Consumer's accuracy reflects the map's reliability from the perspective of a map user. It is calculated by dividing the number of correctly identified pixels in a specific class by the total number of pixels labeled

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

as belonging to that class. This metric indicates the proportion of pixels classified as a certain class on the map that truly belong to that class in reality.

On the other hand, Producer's accuracy represents the map's accuracy from the perspective of the map creator (the "producer"). It is determined by dividing the number of correctly classified pixels for a specific class by the total number of pixels that truly belong to that class. This measure indicates the proportion of pixels in a given class that have been accurately classified. Finally, overall accuracy indicates the proportion of reference data that was correctly classified. It is calculated by dividing the total number of correctly classified pixels by the total number of pixels in the sample as shown in the equation 1 (Nicolau et al, 2024).

$$OverallAccuracy = (TotalPositive + TotalNegative)/Total\ Sample \quad \text{..........} (1)$$

Results and Discussion

This research utilized the Google Earth Engine (GEE) platform for collecting and processing satellite imagery. GEE is a cloud-based platform that provides the capability to process hundreds of satellite imagery data (Phan et al., 2020). Figure 2 illustrates the calculated spectral indices based on the formulas provided in Table 2. All pixels in each figure were scaled to a range of 0 to 1, corresponding to a color gradient from black to white.

Based on the spectral index images, it is observed that some indices, such as MNDWI, NDPonI, EMBI, S2WI, and NDWI, perform well in detecting water bodies compared to other indices. On the other hand, NDVI is effective in classifying vegetation, as vegetation objects produce higher values than non-vegetation objects. Additionally, GEMI produces intriguing results, as it generally yields high values for all cover types.

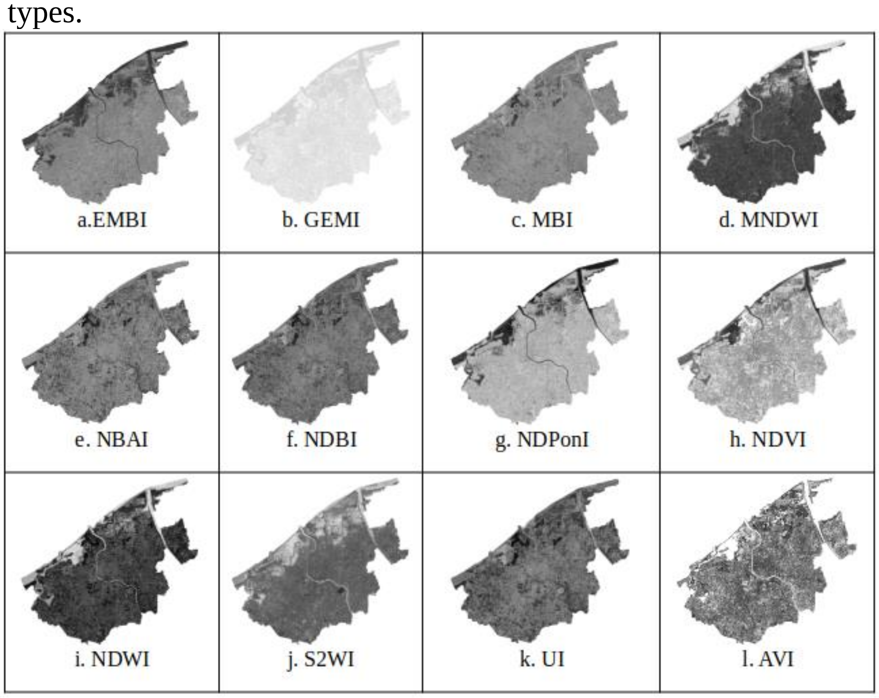

Figure 2. Spectral Indices Images Result

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

Accurately classifying land cover is of utmost importance in land management and environmental monitoring. Sentinel-2 satellite data has emerged as a valuable source of high-resolution data for land cover classification. In utilizing the Sentinel-2 spectral bands, an impressive overall classification accuracy of 87% was achieved. However, it is worth noting that the accuracy level can vary based on various factors, such as the quality of the data, classification method, and the training dataset.

To better understand the classification process, researchers have examined the relative importance of different spectral bands in the Random Forest classification method, as depicted in Figure 3. The results indicate that band B4 played the most crucial role, contributing 11.8% to the classification model. Band B2 was also found to be important, accounting for 10.9% of the model. Additionally, bands B12 and B8 significantly contributed to the classification model, with important roles of 10.5% and 9.7%, respectively. In contrast, the remaining bands had important roles below 9%, with the lowest being band B6, contributing only 6.4% to the model.

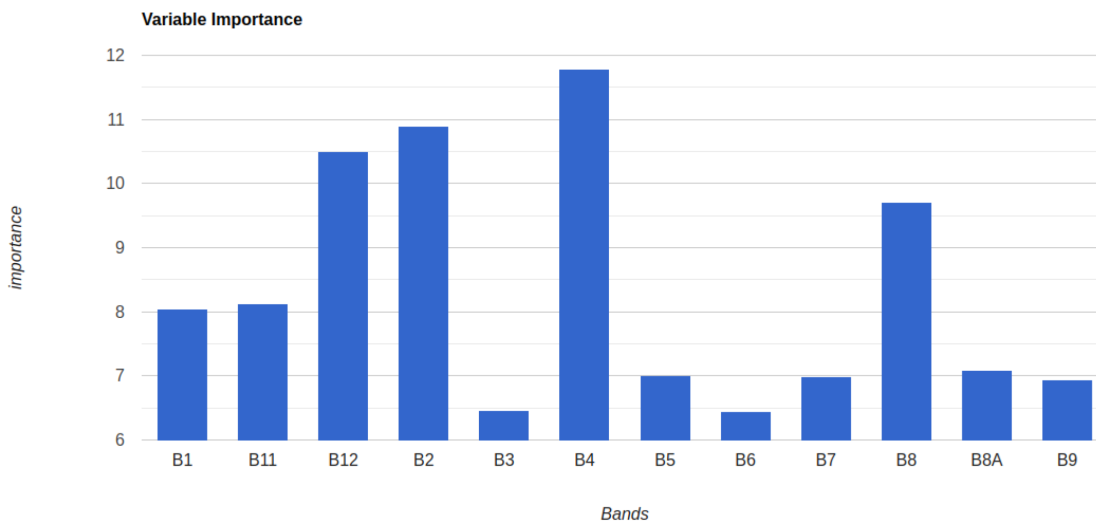

Figure 3. Sentinel-2A Bands Importance in Classification Model

In the Random Forest classification model, all the spectral indices were applied as input variables. Now there are 24 variables that are used to do landcover classification. After training the model, the classification accuracy improved to 90.1%. Moreover, using the confusion matrix as shown in table 3, the classification accuracy can be divided in two groups as consumer accuracy and producer accuracy. The consumer accuracy for water, vegetation, built areas, unbuilt areas and open land respectively 0.87, 1, 0.86, 0.93 and 1. It means for water class a user has 87% chances to get a correct water class. The same explanation goes to other classes. On the other hand the producer accuracy measures the number of correct testing points over all testing points for the class. Producer accuracy for three classes namely water, vegetation and built areas more than 90%. The rest classes which are unbuilt areas and open land have producer accuracy respectively 86% and 50%. The producer accuracy for open land is quite low 50% due to the small number of testing points.

 DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

From 6 testing points only 3 points are classified correctly as open land. Other testing points are misclassified as built area and water.

Table 3. Classification Accuracy Using Sentinel-2 Bands and 12 Spectral Indices

		Classification						
	Class	Water	Vegetation	Built Area	Unbuilt Area	Open Land	Total	Producer Accuracy
Validation	Water	46	0	0	1	0	47	0.98
	Vegetation	0	18	0	2	0	20	0.90
	Built Area	3	0	36	0	0	39	0.92
	Unbuilt Area	2	0	5	43	0	50	0.86
	Open Land	2	0	1	0	3	6	0.50
	Total	53	18	42	46	3	162	
	Consumer Accuracy	0.87	1.00	0.86	0.93	1.00		

Figure 4 shows that B4 still play the highest importance with percentage of 6.6% followed by NBAI with 6%, then NDVI 5.6%, NDVI 5.3%, B12 5.3%, B2 5.2% and B11 5.1%. The rest bands are below than 5%. The lowest importance variables, three from the bottom are: MNDWI, NDPonI and EMBI with respective percentage 2.2%, 2.3% and 2.5%.

After conducting a comprehensive analysis, the results indicate that the Normalized Difference Water Index (NDWI), a commonly used spectral index for identifying water bodies, exhibited a significantly higher level of importance at 4% compared to the Modified Normalized Difference Water Index (MNDWI) at 2.3% and the Simple Ratio Water Index (S2WI) at 3.5%. This outcome suggests that NDWI is an effective spectral index for water body identification when applying in the Random Forest method. Additionally, the analysis revealed that the Normalized Difference Vegetation Index (NDVI), a renowned spectral index for vegetation detection, played a crucial role with a percentage of 5.6%, which surpassed the Advanced Vegetation Index (EVI) at 4.4%. In terms of urban index importance, the Normalized Burn Area Index (NBAI) displayed the highest level of significance at 6%, outperforming other urban indices such as Urban Index (UI) at 4.9%, Normalized Difference Built-Up Index (NDBI) at 3.4%, and Global Environmental Monitoring Index (GEMI) at 4.4%.

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

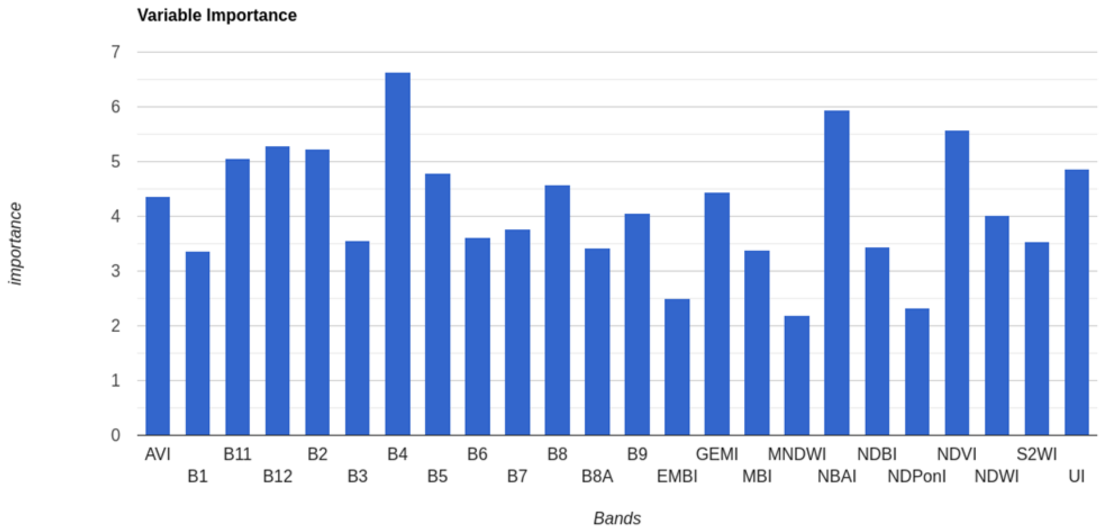

Figure 4. Sentinel-2A Bands Importance in Classification Model

Figure 5 shows the land cover classification result using Sentinel-2 bands and all mentioned spectral indices. The area of each classified land cover type is shown in Table 4. From the table, it can be observed that the overall area of Banda Aceh city is 5997.7 hectares, which differs slightly from the official figure of 6136 hectares. This discrepancy is due to the fact that the pixel resolution of the satellite imagery, which is 10 meters, does not perfectly align with the actual city boundary.

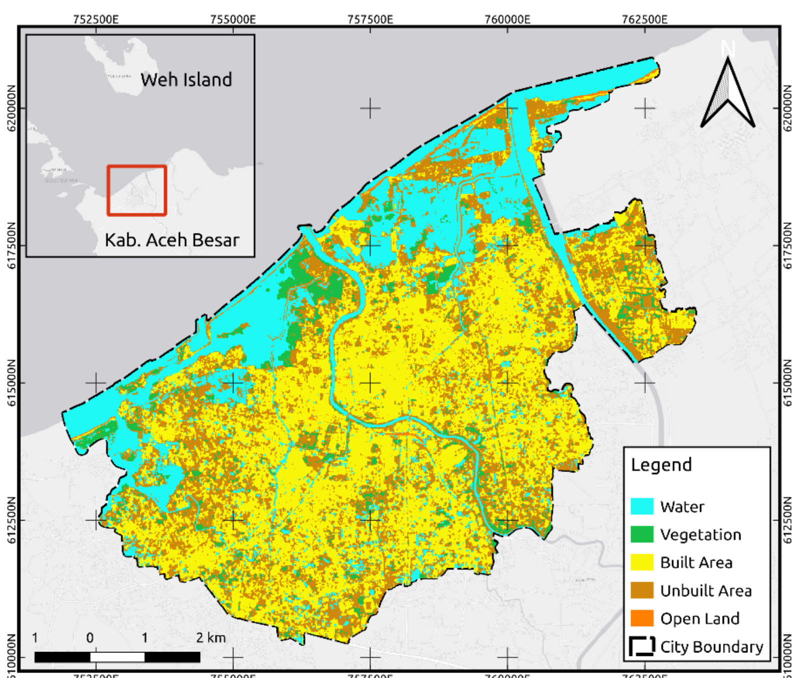

Figure 5. Land Cover Classification Result

Upon comparing the land cover classes between Sentinel-2A bands and Sentinel-2A bands that include all indices, it was discovered that the water body and

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

built area have a very small difference of around 4 hectares. Vegetation and built area increased by 11.7 and 16.4 hectares. Conversely, the open land area decreased from 84.2 hectares to 56.5 hectares.

Table 4. Land Cover Class Area

Land Cover Class	Area (Ha)		
	Sentinel-2A Bands	Sentinel-2A Bands + Indices	Absolute Difference
Water	1360.3	1364.6	4.3
Vegetation	299.0	310.7	11.7
Built Area	2600.6	2617.0	16.4
Unbuilt Area	1653.6	1648.8	4.8
Open Land	84.2	56.5	27.7
Total	**5997.7**	**5997.7**	

Observing the average spectral signature response in figure 6, the spectral indices works as its purpose. For example vegetation indices like NDVI and AVI give strong response for vegetation land cover. Urban indexes like UI, NDBI and NBAI show similar responses where Open land and Built area have strong responses and are quite close to each other. Unfortunately, this similarity is bad for Random Forest input due to its strong correlation that leads to incorrect classification results. Moreover, water indices namely NDWI, S2WI and MNDWI show relatively high response for water objects.

Xu (2006) research points out a potential shortcoming of the Normalized Difference Water Index (NDWI) that should be taken into account when interpreting its classification results. The index's spectral characteristics may have an overlap with those of built-up areas, which can lead to an overestimation of the water areas in the classification. This overestimation can be detrimental to the accuracy of the classification as it can result in misclassifying other features, such as roads, as water bodies. The presence of these misclassified features can significantly reduce the overall precision and reliability of the classification results. Therefore, it is important to take into account this limitation of NDWI when analyzing and interpreting its results.

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

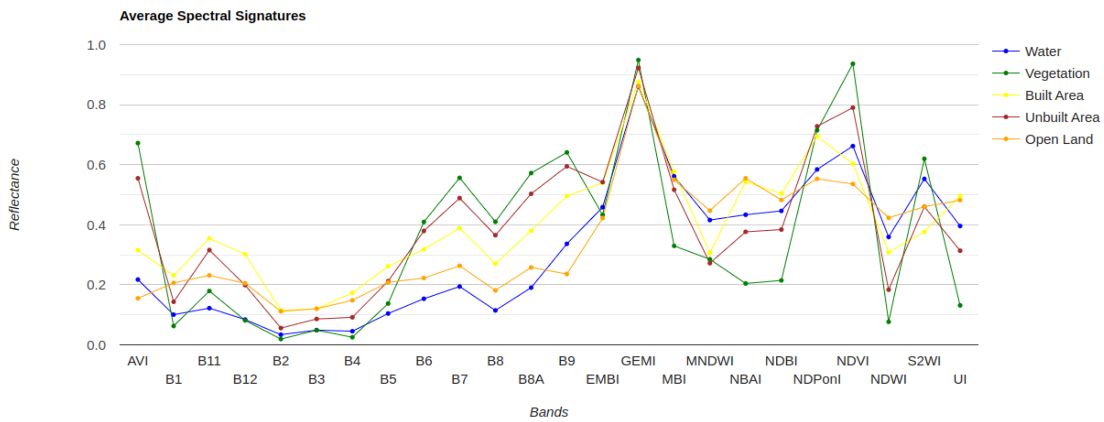

Figure 6. Spectral Signature Response

Conclusion

Accuracy of Land Cover classification using the Random Forest algorithm can be increased by adding more parameters. The results show that the use of spectral indices in random forest methods can slightly improve land cover classification from 87% to 90.1%, indicating that most pixels were correctly classified. Producer's accuracy is highest for water (98%), vegetation (90%), built area (92%), and unbuilt area (86%), reflecting a strong ability to identify these classes correctly. However, open land exhibits a significantly lower producer's accuracy of 50%, indicating frequent misclassification of this class. Consumer's accuracy is perfect for vegetation and open land (100%) and remains high for unbuilt area (93%), water (87%), and built area (86%). Despite these strengths, notable confusion exists between some classes, such as open land and others, as well as water and built area. Moreover, from 12 Bands of the Sentinel-2A satellite imagery and 12 spectral indices used in the algorithm, B4 plays the highest role with 6.6 %, followed by NBAI 6%, NDVI 5.6%, B12 5.3%, B2 5.2% and B11 5.1%. The other bands are below 5% with three the lowest importance are EMBI 2.5%, NDPonI 2.3% and MNDWI 2.2%.

Acknowledgements

The authors would like to acknowledge the invaluable contribution of all colleagues at Mining Engineering Department of Universitas Syiah Kuala who generously gave their time and shared their insights, without whom this research would not have been possible. Last but not least we would like to thank our family and friends for their unwavering support and encouragement during the process of completing this research

 DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

References

BSNI, 2010. Klasifikasi Tutupan Lahan SNI 7645:2010.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recognition Letters 27, 294–300.

Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Lv, J., Yang, K., Fu, J., Qin, X., 2021. An Effective Water Body Extraction Method with New Water Index for Sentinel-2 Imagery. Water 13.

Kawamura, M., Jayamanna, J., Tsujiko, Y., 1996. Relation between Social and Environmental Conditions in Colombo Sri Lanka and the Urban Index Estimated by Satellite Remote Sensing Data. International Archives of the Photogrammetry and Remote Sensing 31, 321–326.

Lacaux, J.P., Tourre, Y.M., Vignolles, C., Ndione, J.A., Lafaye, M., 2007. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. Remote Sensing of Environment 106, 66–74.

Madasa, Akhona, Orimoloye, Israel R., Ololade, Olusola O., 2021. Application of geospatial indices for mapping land cover/use change detection in a mining area. Journal of African Earth Sciences 175.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing 17, 1425–1432.

Mirza Muhammad Waqar, Johum Fatimah Mirza, Rafia Mumtaz, Ejaz Hussain, 2012. Development of New Indices for Extraction of Built-Up Area & Bare Soil from Landsat Data 1.

Nguyen, C., Chidthaisong, A., Kieu Diem, P., Huo, L.-Z., 2021. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. Land 10, 1–18.

Nicolau, A.P., Dyson, K., Saah, D., Clinton, N. (2024). Accuracy Assessment: Quantifying Classification Quality. In: Cardille, J.A., Crowley, M.A., Saah, D., Clinton, N.E. (eds) Cloud-Based Remote Sensing with Google Earth Engine. Springer, Cham.

Phan, T., Kuch, V., Lehnert, L., 2020. Land Cover Classification using Google Earth Engine and Random Forest Classifier-The Role of Image Composition. Remote Sensing 12.

Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. Vegetatio 101, 15–20.

Rikimaru, A., P.S Roy, S Miyatake, 2002. Tropical Forest Cover Density Mapping. International Society for Tropical Ecology 43, 39–47.

Rouse, J. W., Jr., Haas, R. H., Schell, J. A., Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA. Goddard Space Flight Center 3d ERTS-1 Symposium Vol. 1, Sect. A.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing 27, 3025–3033.

DOI : 10.22373/ijes.v2i2.5957
ISSN 3046-885X

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journal of Remote Sensing 24, 583–594.

Zhao, Y., Zhu, Z., 2022. ASI: An artificial surface Index for Landsat 8 imagery. International Journal of Applied Earth Observation and Geoinformation 107, 102703.