The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 6, December 2024
https://doi.org/10.33022/ijcs.v13i6.4510

Comparison of Support Vector Machine and Random Forest Methods on Sentinel-2A Imagery for Land Cover Identification in Banda Aceh City Using Google Earth Engine

Safira[1], Sri Azizah Nazhifah[2*], Muslim Amiren[3], Muhammad Rusdi[4], Nizamuddin[5], Alim Misbullah[6]
safira_20@mhs.usk.ac.id[1], sriazizah07@usk.ac.id[2*], muslim.amiren@usk.ac.id [3], emrusdi@usk.ac.id[4], nizamuddin@usk.ac.id [5], misbullah@usk.ac.id[6]
[1,2,3,4,5,6] Department of Informatics, Syiah Kuala University, Darussalam, 23111, Banda Aceh

Article Information

Received : 20 Nov 2024
Revised : 5 Dec 2024
Accepted : 30 Dec 2024

Keywords

Support Vector Machine, Random Forest, Land Cover, Google Earth Engine, Sentinel 2A,

Abstract

Land cover is a physical feature of the earth that illustrates the relationship between natural processes and social processes. Over time, there has been a lot of land conversion, where initially open land is now built-up land. This is due to the large-scale development in Banda Aceh City. Therefore, this study aims to compare the performance of two classification methods, namely using Support Vector Machine (SVM) and Random Forest in identifying land cover in Banda Aceh City using Sentinel-2A imagery via the Google Earth Engine platform. As for data recording, it starts from January 1 to December 31, 2023. There are 4 classes used in this study, namely vegetation, water bodies, built-up land, and open land. The classification results show that the Support Vector Machine and Random Forest methods have been successfully applied to identifying land cover in Banda Aceh City using Sentinel-2A imagery. The accuracy results show that the Support Vector Machine method has a higher accuracy value of 90.5% compared to the Random Forest method of 85.7%.

A. Introduction

Banda Aceh is one of the cities in Aceh and is the capital of Aceh Province. The city borders the Indian Ocean to the west, the Malacca Strait to the north, Aceh Besar Regency to the east and south. The city of Banda Aceh is of course the center of the economy, education and tourism. Large-scale development has shifted energy from empty land to built-up land. Limited land resources have an impact on land modification which causes changes in land cover conditions. This is because the need for land is inversely proportional to the availability of land. The greater the population, the greater the need for land. The high population and density is one of the factors that puts pressure on land use in the City of Banda Aceh. Therefore, information can be carried out through remote sensing techniques [1].

Remote sensing is the science and art of obtaining information about an object, area, or phenomenon being studied without direct contact with the object [2]. This remote sensing technology has been widely used to provide diverse spatial information on the earth's surface, such as land cover identification [3]. However, so far in processing geospatial data or remote sensing still using manual methods. In this study, remote sensing techniques are used because if done using field surveys, it takes a long time and is also too expensive [4]. This technique can process satellite imagery using the supervised method. The satellite imagery used in this study is Sentinel-2A.

Sentinel-2 is a remote sensing satellite launched in 2015 with 13 spectral bands [5]. This satellite has a spatial resolution of 10 meters for four bands, 20 meters for six other bands, and 60 meters for the remaining three bands [6]. Sentinel-2A provides imagery with high temporal resolution, allowing regular monitoring and detection of changes over time. With good spatial resolution, Sentinel-2A produces imagery with sufficient detail to identify and classify various types of land cover in Banda Aceh City. This satellite is often used for identification and analysis of environmental problems, urban planning, monitoring land cover changes, disaster risk assessments, and other applications [7]. As in one study using a comparison between Sentinel-2A and Landsat-8 imagery to compare the level of accuracy. The results showed that Sentinel-2A imagery provided better accuracy compared to Landsat-8 [8]. The information obtained from this research is also used for other planning and mapping applications [9].

Various studies have been conducted to identify land cover in Banda Aceh City using various methods such as Minimum Distance, Maximum Likelihood [10]. However, this study will use a supervised classification approach with two methods, namely Support Vector Machine (SVM) and Random Forest in 2023. Support Vector Machine (SVM) is a non-parametric supervised machine learning method that can identify and predict patterns based on the training data used [11]. This SVM method is used to find the best hyperplane by maximizing the distance between classes. while Random Forest is able to handle large amounts of training data efficiently. Where later good performance in this classification will result in few errors. This Random Forest method is a predictive tree method, where each tree independently relies on samples from random vectors and has the same distribution for all trees in the forest [12].This study will focus on four classes, namely vegetation, water bodies, open land, and built-up land.

B. Research Method

This study uses two methods, namely Support Vector Machine (SVM) and Random Forest (RF). The purpose of using these two methods is to compare which method produces the best accuracy. Based on previous studies, several studies have shown that the Random Forest method produces the highest accuracy, while other studies have shown that Support Vector Machine provides high accuracy results. This study aims to test the application of the SVM and RF methods in identifying land cover in Banda Aceh City, as well as to compare the accuracy of the two methods in identifying land cover using Sentinel-2A Imagery. Determination of land cover is carried out using the 2 methods on Sentinel-2A Imagery using the GEE platform, through several stages such as handling cloud cover and cloud masking, so that the satellite imagery used is free from clouds or has a minimal percentage of clouds. Furthermore, the SVM and Random Forest algorithms are applied to four classes of land cover, namely vegetation, water bodies, open land, and built-up land, with the number of samples taken for each class as many as 100 samples. The final results of this classification are then exported and saved for further analysis using ArcGIS. Accuracy testing is carried out using a confusion matrix and also Kappa accuracy and visual verification through existing field images. The workflow in this study can be seen in Figure 1.

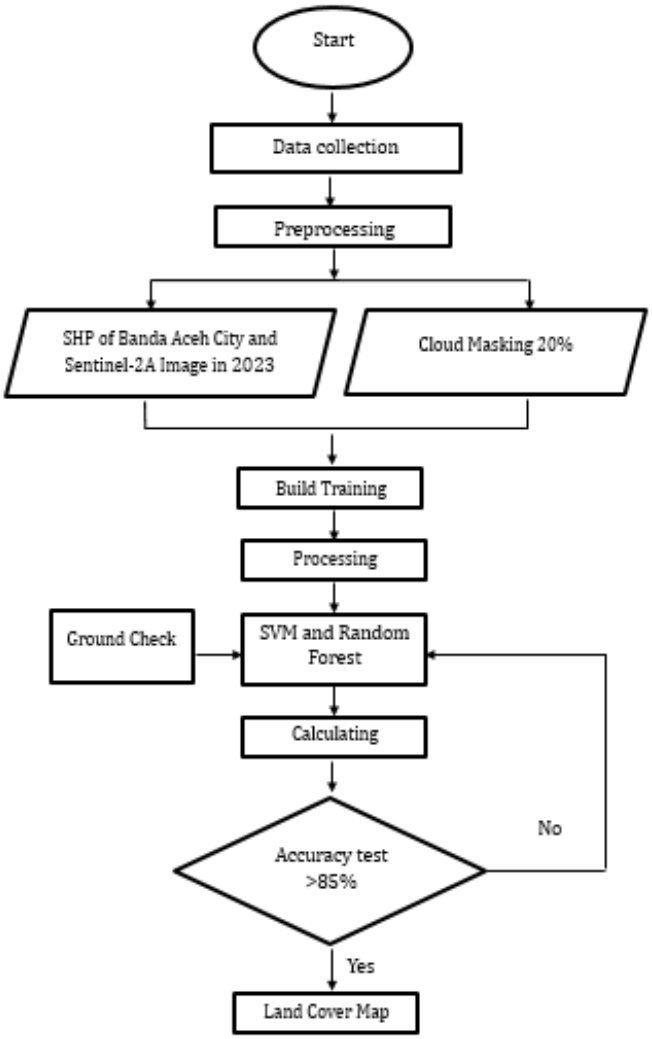

Figure 1. Research Flow Diagram

- Identification and Formulation of Problems. At this stage is the initial step of this research by involving the identification and understanding of a problem that will be raised in this research. Identification of the problem here will involve a deep understanding of the theme or title raised in this research. After that, the problem requires several formulations of specific, relevant, and motivating research questions. This process will ensure that the research has a clear direction and flow in solving several problems.
- Literature Study. At this stage is an important stage in this research because it involves a thorough understanding of the literature, be it books, journals, and other sources related to the research topic. The purpose of this literature study is to understand changes in knowledge and research in a particular field, identify knowledge gaps, and build a conceptual foundation to guide further research. This process allows researchers to evaluate the research methods used in previous studies, identify important findings, and determine the relevance of the literature to the problem being studied.
- Data collection. The initial step in this study is to enter data by accessing Google Earth Engine. Furthermore, data in the form of administrative boundaries of Banda Aceh City in SHP format is entered by clicking on the asset manager located on the left on the GEE platform. After that, the Sentinel-2A image is input by selecting the "browse catalog" option and selecting the Sentinel type, namely COPERNICUS/S2 SR HARMONIZED or entered through the Image Collection call available on GEE. The imagery used in this study starts from January 1 to December 31, 2023
- Cloud Cover. This process involves a method called cloud cover. This method is better if the cloud cover is less than 20%. Several methods can be used in GEE to minimize cloud cover in the image. The first method can be done with a cloud cover filter, which allows us to sort Sentinel-2A and Landsat-8 images from a certain period of time with the desired cloud cover. The weakness of this method is that the image cannot be completely free of clouds. The cloud cover filter helps select only satellite images with the least cloud cover, so that other processes. The mask filter is used to combine several images that are illuminated in one recording session to obtain an image with a clearer and cleaner cloud layer. The masking process works by replacing the image pixels covered by clouds with pixels of the masked image using BQA pixels from Sentinel-2A and QA60 images, which contain information on the percentage of cloud cover per pixel.
- Build Training. At this stage of building training, first ensure that the variables that have been created previously can be imported. Then in the variable will store a list of previously determined band names, namely B1 to B7 for feature extraction on Sentinel images. Then sample regions are used to take pixel samples at the selected location, namely the city of Banda Aceh which is defined by the Area Of Interest (AOI) variable. After the process is carried out, the training variable will contain a dataset containing pixel values from the selected band and labels from the properties of each class that has been created, namely vegetation, water bodies, built-up land and open land.

- Support Vector Machine and Random Forest Classification. For images that have gone through the 2 processes mentioned to remove cloud cover such as cloud cover and cloud masking, then the next step is classification using the Support Vector Machine (SVM) and Random Forest methods. In this process, for each class that has been created with 100 samples for each class, all classes will be merged or combined so that they will display results according to the specified class. For the selection of sample points, accuracy is needed to minimize errors in classification.
- Export Data. After getting the results of the classification, the next step is to export the data. Data processed in GEE can be exported to data formats such as images, tables, maps, and videos. Where the exported data can be directly saved to a cloud storage drive.
- Accuracy Test. In the accuracy assessment, it is used to see how much error occurs in the classification by sampling so that the percentage of map accuracy can be determined. This study tests the level of image accuracy in supervised classification. The accuracy of the test carried out in this study is the error matrix or confusion matrix. In this contingency matrix, it can also calculate the magnitude of the producer's precision (producer's accuracy) and user accuracy (user accuracy) used by each class. For the accuracy testing stage, previously the classification results that had been exported were entered into ArcGIS by randomly marking 100 points using the random point feature. Then later the ground check process will be carried out via Google Earth Pro by marking several locations randomly in the class that has been created. The formula can be seen in the following equation:

$$\text{User's Accuracy} : \frac{X_{ii}}{X_{i+}}\ x\ 100\% \quad (1)$$

$$\text{Producer's Accuracy} : \frac{X_{ii}}{X_{+i}}\ x\ 100\% \quad (2)$$

$$\text{Overall Accuracy} : \frac{(X_{11}+X_{22}+X_{33}\ldots+X_{ii})}{X_{total}}\ x\ 100\% \quad (3)$$

Description:
Xii = Diagonal value of row i and column i of the partition matrix
Xi+ = Number of pixels in row i
X+i = Number of pixels in column i

- Layouting. The last step taken in this research is to display the results in the form of a map layout. The layout contains standard elements that exist on the map and also adds the geographical location of Banda Aceh geographically, area, type of map used, color information, and classes that have been created.

C. Results and Discussion

When classifying land cover using 2 methods, namely Support Vector Machine and Random Forest, there were several trials conducted to produce good classification. There were 11 stages of testing and the results were not good because when the sample selection was carried out in each class it was not careful so that for each class when combined into another pixel, the results were more dominant towards one class. Then another factor that affects classification errors is the lack of training data, because the more training data the better the results obtained. Furthermore, another factor that affects classification errors is the difference in parameters or classification methods between SVM and Random Forest which can cause different classification results so that misclassification occurs. In this study, good SVM classification results were found in SVM without using parameters, because when the value of the parameter is used it is difficult to find the right value to get the best performance. Here when random values are used for each parameter must have many guesses in order to get the best performance. Furthermore, for the Random Forest method, good results were found in 10 decision tree values, because the more trees are built, the better the performance results. For good classification results, there were in the 12th trial. Furthermore, to display the overall results of the classification process, it can be displayed by adding layers and colors according to the classes that have been created.

Classification Results

After the 12th stage was carried out, good results were obtained, resulting in 4 classes as previously determined. The classification results can be seen in Figure 2.

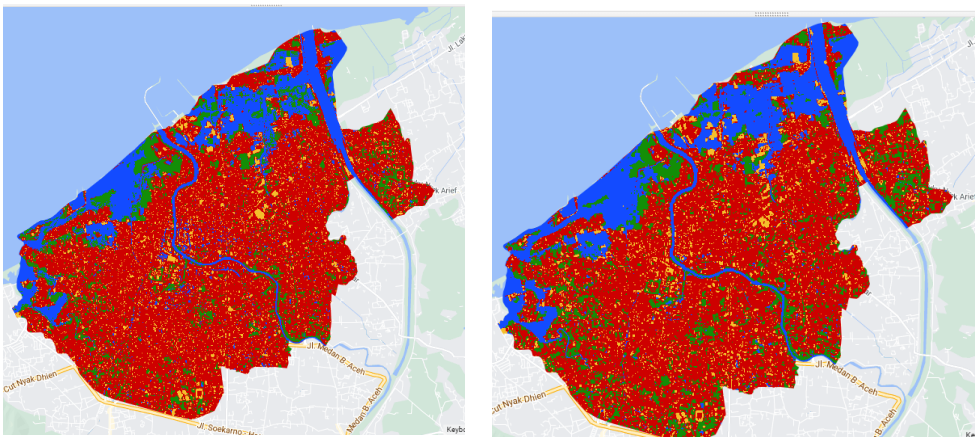

Figure 2. Random Forest and SVM classification results

In Figure 2 above is the result of classification using Google Earth Engine which consists of 4 classes, namely vegetation, water bodies, built-up land, and open land using the Random Forest and Support Vector Machine methods. Where the green color indicates the vegetation class, blue as water bodies, red as built-up land, and yellow as open land. These results were obtained after going through the process mentioned earlier. Furthermore, the classification results are exported in Geo TIFF format.

Accuracy Testing

At this stage of accuracy testing is carried out to obtain the level of accuracy of the results obtained from the Support Vector Machine and Random Forest methods. The results of this accuracy will be displayed in the form of a confusion matrix table by calculating user's accuracy, producer's accuracy, and overall accuracy. To calculate producer's accuracy is obtained by dividing the pixel value that is counted as correct in one class by the number of pixels in the column, while for user's accuracy is divided by the number of pixels in the row. The results of random points can be seen in Figure 3.

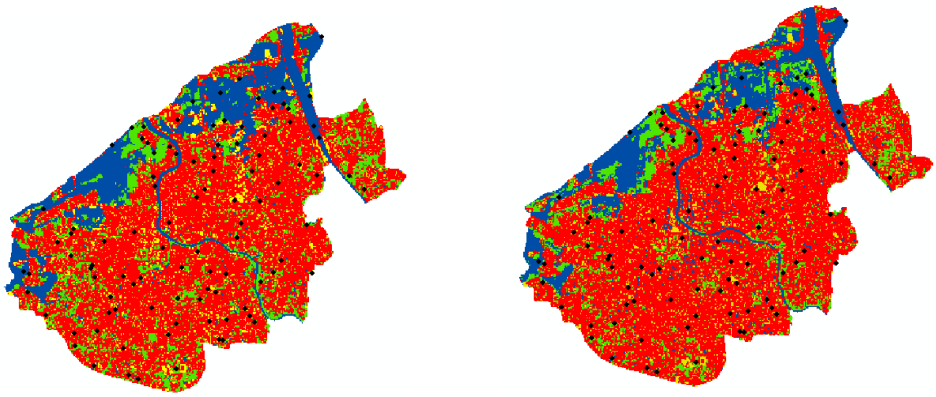

Figure 3. Random Point SVM and Random Forest Results

The image above shows the results of random points performed on ArcGIS software. Before the process is carried out, the initial steps taken involve calculating statistics and determining colors according to the predetermined class. Then, the background is removed from the results by extracting using the 'extract by mask' method. Furthermore, random points are performed with a total of 100 points. Next, a ground check will be carried out via Google Earth Pro, where the results of the random points are converted into KML. The results of the ground check can be seen in Figure 4.

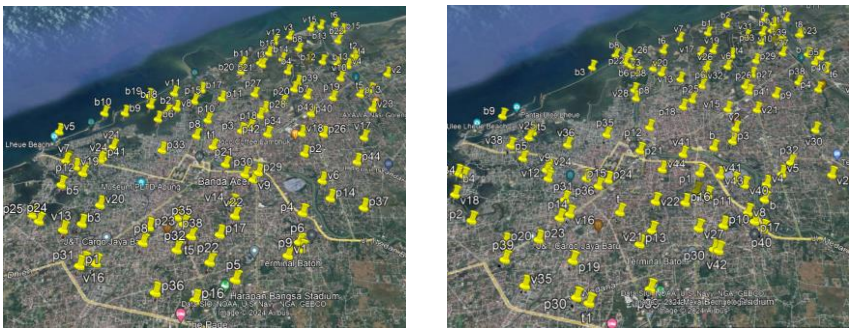

Figure 4. Ground Check Results on Google Earth Pro Using SVM and Random Forest

The image above shows the results of a ground check via Google Earth Pro using the SVM and Random Forest methods. Both methods have been used for 100 random points, then converted into KML format so that they can be entered into the Google Earth Pro application. Each point is given coordinates by adding them to the placemark, and given a code, for example v1, p1, b1, t1, and so on. This is done to make it easier when checking one by one via ArcGIS. Furthermore, the results from Google Earth Pro are opened via ArcGIS and converted into SHP format for checking. The results can be seen in Figure 5

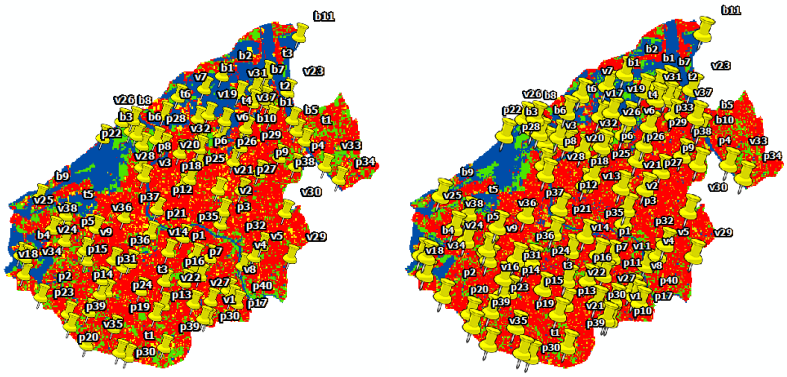

Figure 5. Conversion Results from KML to SHP

In the image above, the results of the KML conversion into SHP format are visible. After these coordinate points are entered into ArcGIS, a checking process is carried out for 100 points that have been previously randomized. This checking process is carried out one by one, for example for the vegetation class, then the suitability is seen. If the result is correct, then a value of 1 to 4 classes that have been created is given, so that the final result will be calculated as overall accuracy, user's accuracy, and producer's accuracy. For the accuracy results, they can be seen in Tables 1 and 2.

Table 1. Support Vector Machine Confusion Matrix

SVM Classification Results	***Ground Truth***				**Total**
	Vegetation	**Water Body**	**Built-up Land**	**Open Land**	
Vegetation	**27**	0	1	0	28
Water Body	0	**13**	1	0	14
Built-up Land	1	2	**48**	1	52
Open Land	0	0	0	**6**	6
Total	28	15	50	7	**100**

Table 2. Random Forest Confusion Matrix

Random Forest Classification Results	***Ground Truth***				**Total**
	Vegetation	**Water Body**	**Built-up Land**	**Open Land**	
Vegetation	**24**	2	2	0	28
Water Body	0	**13**	1	0	14
Built-up Land	3	1	**48**	0	52
Open Land	0	0	0	**6**	6
Total	27	16	51	6	**100**

Based on Tables 1 and 2 above, it can be seen that the user accuracy value, producer accuracy, and overall accuracy for the Support Vector Machine method are 90.5%, which is a higher accuracy compared to Random Forest which only reaches 85.7%. The SVM method is superior because in this land cover classification it uses the best hyperplane as a separator between classes, so this method can read classes well so that it produces high accuracy. Both of these accuracy values have met the minimum standards set by the USGS, which is ≥ 85%. This is in line with previous research which uses the K-Means Clustering method with a division of 3 classes, namely vegetation, non-vegetation, and water bodies with an accuracy value reaching 90% accuracy with Sentinel-2A Imagery [13]. However, the weakness of this method is that users cannot label classes directly so that use on the GEE platform will group classes based on the closest value. Therefore, better results can be obtained by using supervision methods such as Support Vector Machine which have been studied previously.

In the Support Vector Machine method, there was 1 misclassification in the vegetation row which was misclassified as built-up land from a total of 28 samples. Then, there was a misclassification in the water body, where 1 point was misclassified as built-up land from a total of 14 samples. In the built-up land class, there were 4 misclassifications, with 1 point misclassified as vegetation, 2 points as water bodies, and 1 point as open land. Meanwhile, in the Random Forest method there were 4 misclassifications, where 2 points were identified as built-up land and 2 points as water bodies from a total of 28 samples. In the water body class, there was 1 misclassification identified as built-up land from a total of 14 samples. Then in the built-up land class there were 4 misclassifications, with 3 points identified as vegetation and 1 point identified as water bodies from a total of 52 samples. Based on the results of the classification carried out, there were 4 classes, namely vegetation, water bodies, built-up land and open land, so the area value for each class was obtained using 2 methods, namely Support Vector Machine and Random Forest. The area can be seen in Tables 3 and 4.

Table 3. Area for the Support Vector Machine method

No	Class name	Area (Hectares)
1	Vegetation	82.686 Ha
2	Water Body	99.352 Ha
3	Built-up Land	338.250 Ha
4	Open Land	39.515 Ha

Table 4. Area size for Random Forest method

No	Class name	Area (Hectares)
1	Vegetation	67.342 Ha
2	Water Body	88.223 Ha
3	Built-up Land	365.754 Ha
4	Open Land	38.484 Ha

In tables 3 and 4 above, the total area of each class is explained. Of the four classes, it can be seen that Banda Aceh City is more dominated by built-up land. The

SVM method produces an area of 338,250 hectares, while the Random Forest method produces an area of 365,754 hectares. In addition, the area for the open land class is very limited, indicating that much open land in Banda Aceh City has turned into built-up land, this is due to the increase in population and also urban development. Then to see how accurate the classification results obtained, then a field survey was conducted to see the suitability between the results using GEE with the actual results. For the results, they can be seen in Tables 5, 6, 7, and 8.

Table 5. Results of Ground Check Field Land Cover Classification of Banda Aceh City (Vegetation)

No	Field Survey	Classification in GEE (SVM)	Classification in GEE (Random Forest)	Information
1.	Vegetation	Vegetation	Vegetation	True
2.	Vegetation	Vegetation	Vegetation	True
3.	Vegetation	Vegetation	Water Body	False
4.	Built-up Land	Vegetation	Vegetation	False

Table 6. Results of Ground Check Field Land Cover Classification of Banda Aceh City (Water Body Class)

No	Field Survey	Classification in GEE (SVM)	Classification in GEE (Random Forest)	Information
1.	Water Body	Water Body	Water Body	True
2.	Water Body	Water Body	Water Body	True
3.	Water Body	Water Body	Water Body	True
4.	Water Body	Water Body	Water Body	True

Table 7. Results of Ground Check Field Land Cover Classification of Banda Aceh City (Built-up Land Class)

No	Field Survey	Classification in GEE (SVM)	Classification in GEE (Random Forest)	Information
1.	Built-up Land	Built-up Land	Built-up Land	True
2.	Built-up Land	Built-up Land	Built-up Land	True
3.	Built-up Land	Built-up Land	Built-up Land	True
4.	Open Land	Built-up Land	Open Land	False

Table 8. Results of Ground Check Field Classification of Land Cover in Banda Aceh City (Open Land)

No	Field Survey	Classification in GEE (SVM)	Classification in GEE (Random Forest)	Information

1.	Open Land	Open Land	Open Land	True
2.	Open Land	Open Land	Open Land	True
3.	Open Land	Built-up Land	Built-up Land	False
4.	Vegetation	Open Land	Built-up Land	False

The table above shows the results of the field survey with the results carried out on GEE. There are 4 classification errors, where in the vegetation class, 1 point is identified as a water body and 1 point is identified as built-up land, then 1 built-up land class is identified as open land, while 1 open class is identified as vegetation from a total of 16 test points. Furthermore, the differences in the results of the Support Vector Machine and Random Forest can be seen in Figure 6.

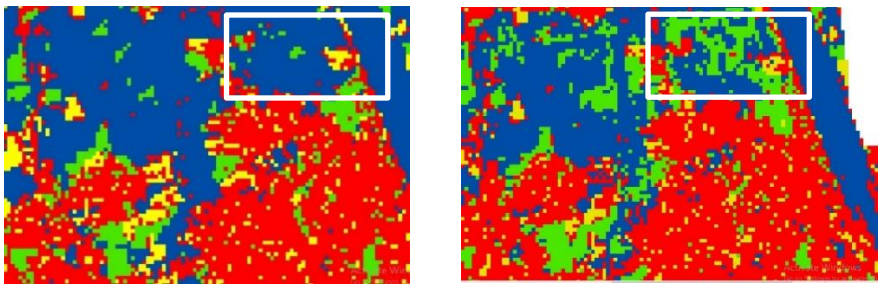

Figure 6. Differences in Land Cover Classification Results using Support Vector Machine and Random Forest

In the image above, the difference in land cover classification results can be seen, where there are several differences in pixel colors, where the area that should be a water body class is included in the vegetation class, as well as the other classes. From the image, it can be seen that the most accurate and appropriate results are in the Support Vector Machine method.

Land Cover Map

Classification Results

Land cover map built on ArcGIS software using a scale of 1:61,000 and the WGS 1984 coordinate system. The creation of this land cover map is based on standards SNI 7645 that meet cartographic criteria. For the land cover map of Banda Aceh city processed using the Support Vector Machine and Random Forest methods, it can be seen in Figures 7 and 8.

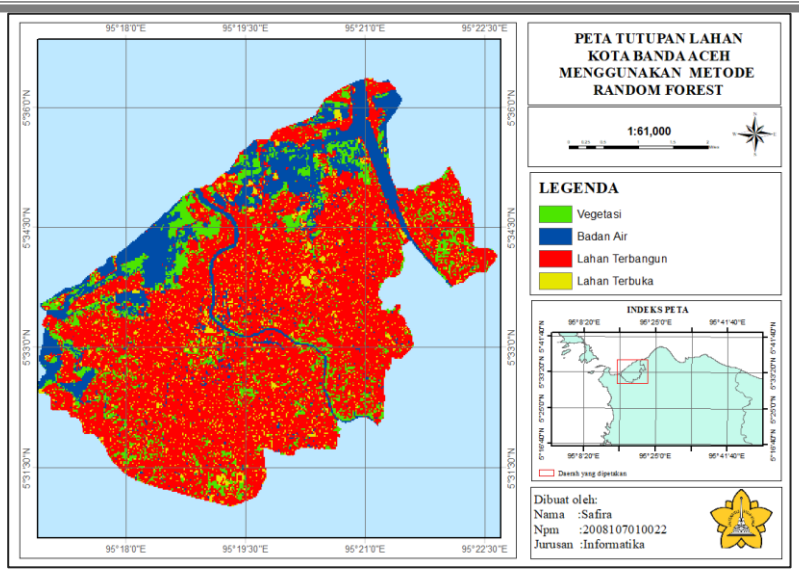

Figure 7. Land Cover Map of Banda Aceh City Using the Random Forest Method

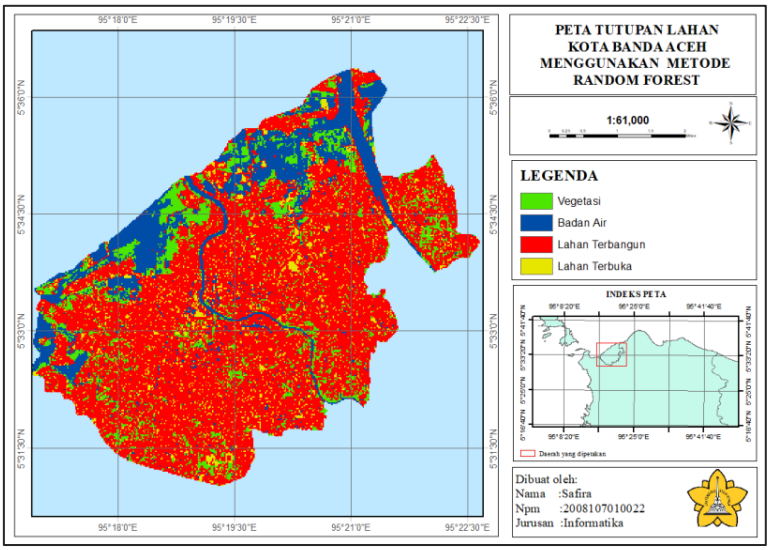

Figure 8. Land Cover Map of Banda Aceh City Using the Support Vector Machine Method

D. Conclusion

Based on the research results explained above, it can be concluded that the Support Vector Machine and Random Forest methods were successfully applied to classify land cover in Banda Aceh City into 4 classes, namely vegetation, water bodies, built-up land, and open land. The accuracy test carried out on the classification results with 100 test points on Sentinel-2A images using the Random Forest method produced an overall accuracy value of 85.7%, recorded lower than the classification results using the Support Vector Machine method with an overall accuracy value of 90.5%. With both of these values meeting the minimum interpretation of the level of accuracy set by the USGS, the results of this study are included in the feasible category and can be used for other purposes.

E. References

[1] M.S. Rumetna, E. Sediyono, and K.D. Hartomo, "Analisis Perubahan Tata Guna Lahan di Kabupaten Bantul Menggunakan Metode *Global Moran's* I," *J. Buana Informatika,* vol 8, no. 4, pp. 225-234, Sep 2017, doi: 10.24002/jbi.v8i4.1446.

[2] C. Bishop,B. Rivard, C. de Souza Filho and F. Van Der Meer, "Geological remote sensing," *International journal of applied earth observation and geoinformation*, vol 64, no.8, pp. 267-274, Dec 2018, doi: 10.1016/j.jag.2017.08.005.

[3] S.S. Rwanga, and J.M, "Accuracy assessment of land use/land cover classification using remote sensing and GIS," *International Journal of Geosciences*, vol 8, no 4, p. 611, Dec 2017, doi: 10.35940/ijeat.A1011.1291S319.

[4] J.W. Merchant, and S. Narumalani, "Integrating remote sensing and geographic information systems," *The SAGE handbook of remote sensing*, pp. 257-268, Jan 2009, doi 10.4135/9780857021052.n18.

[5] D. Phiri, M, Simwanda, S. Salekin, V.R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol 12, no 14, p. 2291,Jul 2020, doi: 10.3390/rs12142291.

[6] Li. S. Ganguly, S. Dungan, J. L. Wang, W, R.R Nemani, R. R," Sentinel-2 MSI radiometric characterization and cross-calibration with Landsat-8 OLI," *Advances in Remote Sensing*, vol 6, no 2, p. 147, June 2017, doi: 10.4236/ars.2017.62011.

[7] M. Jannah, A. Rahmat, A, W.S. Ramadhani, and N.F. Istiawati," Analisis Tingkat Kehijauan Lahan Daerah Sukarame, Bandar Lampung dengan Metode NDVI Berdasarkan Citra Landsat Tahun 2000 dan 2020," *Jurnal Plano Buana*, vol 2, no 1, pp. 39-47, Nov 2021, doi: 10.36456/jpb.v2i1.3494

[8] D. Duryat, and A. Darmawan," Efisiensi penggunaan citra multisensor untuk pemetaan tutupan lahan," *Jurnal Sylva Lestari*, vol 7, no 3, pp. 342-349, sept 2019, doi: 10.23960/jsl37342-349.

[9] N. Oktaviani, and H.A Kusuma," Pengenalan citra satelit sentinel-2 untuk pemetaan kelautan," *Oseana*, vol 42, no 3, pp. 40-45, July 2017, doi: 10.23960/jsl37342-349.

[10] E. Artika, A. Darmawan, and R. Hilmanto," Maximum Likelihood Clasification (MLC) dan Object Oriented Classification (OOC) dalam Pemetaan Tutupan Mangrove di Kabupaten Lampung Selatan," *Jurnal Hutan Tropis*, vol 7, no 4, pp. 267-275, Nov 2019, doi: 10.20527/jht.v7i3.7513.

[11] S. Firmansyah, J.L. Gaol, and S.B. Susilo," Perbandingan klasifikasi SVM dan Decision Tree untuk pemetaan mangrove berbasis objek menggunakan citra satelit Sentinel-2B di Gili Sulat, Lombok Timur," *Journal of Natural Resources and Environmental Management*, vol. 9, no 3, pp 746-757, Sept 2019, doi: 10.29244/jpsl.9.3.

[12] L. Breiman, Leo,"Random forests," *Machine learning*, 45, vol. 45, pp. 4-32, Oct 2001, doi: 10.1023/A:1010933404324.

[13] D. Rahmadani, "K-Means Clustering on Landsat-8 and Sentinel-2A Images for Land Cover Identification in Banda Aceh City," thesis, Dept. of Informatics, Syiah Kuala University, Banda Aceh, 2023.