IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- ____
Vol 4 No 1 | June 2023

COMPARISON OF RANDOM FOREST AND MAXIMUM LIKELIHOOD CLASSIFICATION METHODS FOR LAND COVER IN LANDSAT 9 IMAGES IN LUBUK KILANGAN DISTRICT

Fitri Hayati[1], Febriandi[2], Ernawati[2], Sri Kandi Putri[2]

[1]Student of the D3 Remote Sensing Technology Study Program, Universitas Negeri Padang, [2]Lecturer Study Program D3 Remote Sensing Technology, Universitas Negeri Padang
e-mail: fitrihayatinasution@gmail.com

ABSTRACT: Information on land cover is needed in various sectors including management and resources which can be obtained through data processing using remote sensing satellite imagery. This research was conducted in Lubuk Kilangan District using Landsat 9 imagery, with the aim of (1) knowing the land cover classification using the random forest method, (2) knowing the land cover classification using the maximum likelihood classification method, and (3) knowing the best method for obtaining land cover information based on the accuracy value between the random forest method and the maximum likelihood classification. The method used is a comparative quantitative method by comparing the random forest method and the maximum likelihood classification of land cover in Lubuk Kilangan District. This study performs classification accuracy test calculations using Kappa with the help of a confusion matrix. The results of the study obtained 13 land cover classes from were found from taking training samples showing (1) the random forest land cover classification method was able to classify images properly. This is proven by findings in the field where 86% of pixels are classified correctly. Meanwhile, (2) the maximum likelihood classification method of land cover classification is not able to classify images properly. This is proven by findings in the field where 55% of pixels are classified correctly. (3) the Kappa accuracy value found for the random forest method is 0.81, while the maximum likelihood classification method is 0.51. This shows that the random forest method is better at obtaining information on the land cover than the maximum likelihood classification method.

Keywords: Land cover, Landsat 9, Random Forest, Maximum Likelihood Classification

1. INTRODUCTION

Land cover is important in environmental studies and natural resource management (Abdullah et al, 2019). Land cover information provides data needed in various sectors including resource management and monitoring (Beaubien et al, 1999). One method that is often used to obtain land cover information is mapping using satellite imagery in remote sensing (Zhang, Y., & Wang, Y. 2018).

Remote sensing is a data source for producing land cover data that can describe the appearance of the earth's surface in a spatial, consistent, and high-coherence manner and is available at various scales both temporally and spatially (Foody, 2001). Remote sensing has become an important and effective means of monitoring land cover with its ability to provide information about spatial variations on the earth's surface quickly, comprehensively, accurately and easily (Gong et al, 2013).

The classification method is one of the methods used to obtain land cover information from satellite imagery. Random forest and maximum likelihood classification methods are often used in land cover classification, but they have different advantages and disadvantages. Comparing the two methods can help in understanding the performance of each method and help in choosing which method is right for a particular situation. The maximum likelihood classification determines the class with the highest probability of a given feature. The advantages of the maximum likelihood classification are its lower dependence on the amount of test data and its higher dependence on the level of information in the data. The disadvantage of the maximum likelihood classification is that assumptions about the distribution of features may be invalid, and this method is very susceptible to outliers (Bishop, 2017). Meanwhile, the random forest method is an ensemble-based classification method that uses several decision trees to make classification decisions. The random forest combines the results of several decision trees to determine the best class in terms of the average or majority divisor.

IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- ____
Vol 4 No 1 | June 2023

The advantage of the random forest is that it does not rely too much on assumptions about the distribution of features and is very good at dealing with outliers. The drawback of the random forest is that it takes longer to train the model compared to the maximum likelihood classification (James et al, 2013). the random forest method is an ensemble-based classification method that uses several decision trees to make classification decisions. The random forest combines the results of several decision trees to determine the best class in terms of the average or majority divisor. The advantage of the random forest is that it does not rely too much on assumptions about the distribution of features and is very good at dealing with outliers. The drawback of the random forest is that it takes longer to train the model compared to the maximum likelihood classification (James et al, 2013). the random forest method is an ensemble-based classification method that uses several decision trees to make classification decisions. The random forest combines the results of several decision trees to determine the best class in terms of the average or majority divisor.

The advantage of the random forest is that it does not rely too much on assumptions about the distribution of features and is very good at dealing with outliers. The drawback of the random forest is that it takes longer to train the model compared to the maximum likelihood classification (James et al, 2013). The advantage of the random forest is that it does not rely too much on assumptions about the distribution of features and is very good at dealing with outliers. The drawback of the random forest is that it takes longer to train the model compared to the maximum likelihood classification (James et al, 2013). The advantage of the random forest is that it does not rely too much on assumptions about the distribution of features and is very good at dealing with outliers. The drawback of the random forest is that it takes longer to train the model compared to the maximum likelihood classification (James et al, 2013).

This study compares the random forest method and the maximum likelihood classification for land cover using Landsat 9 imagery. Then, the results of the two methods used are compared to find out which method has a better performance by knowing the level of accuracy of the two methods.

2. RESEARCH METHODS

The type of research used is quantitative research with a comparative approach. Comparative quantitative methods are used to compare two or more variables or groups. The results of comparative quantitative research can be expressed in the form of numbers or percentages which can then be compared to find out the differences or similarities between two or more variables or groups. The research flowchart is shown in Figure 1.

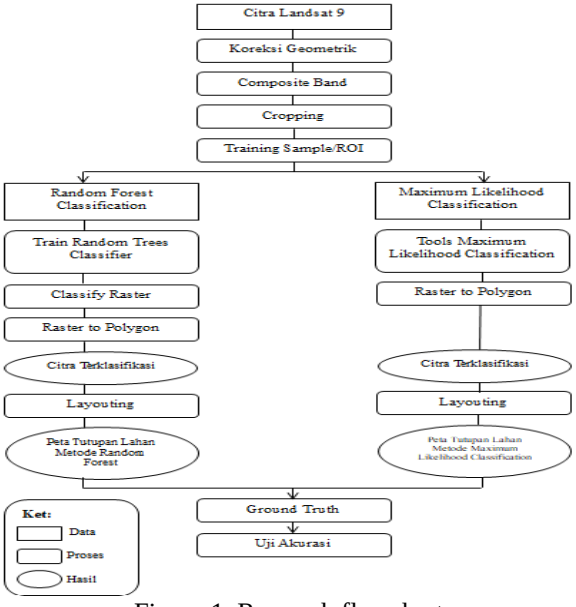

Figure 1. Research flowchart

2.1 Identification of Land Cover on Landsat 9 Imagery

Identifying land cover in Landsat 9 imagery can utilize the 6-5-4 band combination for vegetation analysis because this band combination takes advantage of differences in light absorption by chlorophyll in leaves at different wavelengths. Types of land cover can be analyzed by interpreting the resulting image based on its interpretation

IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- ____
Vol 4 No 1 | June 2023

elements (hue/color, size, shape, texture, pattern, height, shadow, site, association) and determining the land cover class refers to the land cover class according to SNI-7645 of 2010 .

2.2 Image Classification

2.2.1 Random forest method

The random forest method produces many trees that are used as the basis for a majority vote. The majority vote is used to determine the label class in the output. When inputting data, the random forest will create a subset of the data for as many iterations as it does. The process of image classification using the random forest method is carried out by using training samples or regions of interest (ROI) that have been created, and then running with tools to train random trees in the form of raster data which will later be converted into vectors.

2.2.2 Maximum likelihood classification method

The maximum likelihood classification method evaluates quantitatively the variance as well as the correlation of categorical spectral response patterns when classifying unknown pixels. The process of image classification using the maximum likelihood classification method is carried out using training samples or regions of interest (ROI) that have been made, and then running with the maximum likelihood classification tools.

3. RESULTS AND DISCUSSION

3.1 Research result

3.1.1 Classification of land cover random forest method

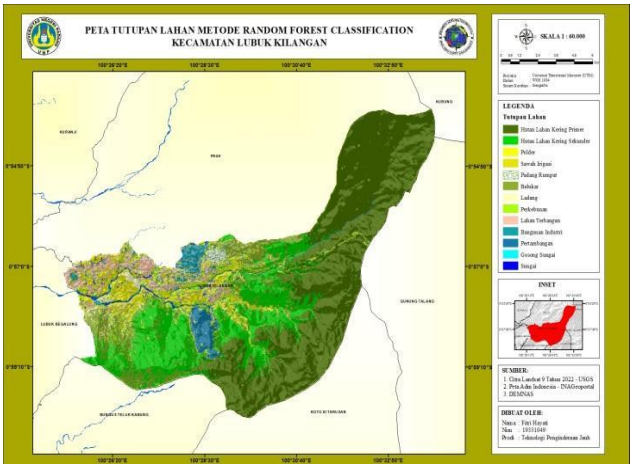

Figure 2. Land cover map using the random forest method

Table 1 The results of land cover using the random forest method

No	Land Cover	Number of Pixels	Area (Ha)	Percentage (%)
1	Primary Dryland Forest	496	3886.61	47
2	Secondary Dryland Forest	1490	1644.61	20
3	polder	2641	303.65	4
4	Irrigation Field	2255	415.99	5
5	Meadow	1386	155.02	2
6	thicket	2399	743.99	9
7	Field	938	126.76	1
8	Plantation	1649	195.33	2
9	Built-up Land	818	368.45	4
10	Industrial Buildings	1218	154.99	2
11	Mining	861	221.54	3
12	Gosong River	314	23.06	0
13	River	388	46.11	1
	Amount	16,853	8286.11	100

Source: Data processing, 2022

IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- _____
Vol 4 No 1 | June 2023

3.1.2 Land cover classification using the maximum likelihood classification method

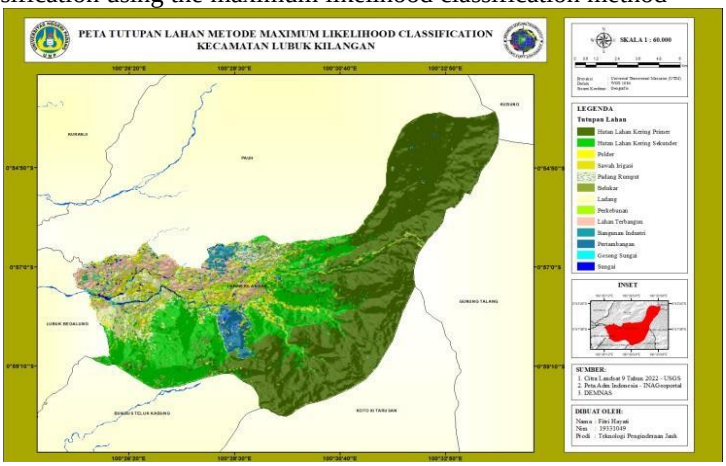

Figure 3. Land cover map using the maximum likelihood classification method

Table 2 The results of the land cover maximum likelihood classification method

No	Land Cover	Number of Pixels	Area (Ha)	Percentage (%)
1	Primary Dryland Forest	538	4173.05	50
2	Secondary Dryland Forest	1079	1752.01	21
3	polder	1637	241.99	3
4	Irrigation Field	1950	326.66	4
5	Meadow	623	169.11	2
6	thicket	1297	147.19	2
7	Field	994	310.27	4
8	Plantation	1697	254.11	3
9	Built-up Land	688	458.23	5
10	Industrial Buildings	398	91.27	1
11	Mining	584	213.54	3
12	Gosong River	1185	90.56	1
13	River	378	58.16	1
	Amount	13,048	8286.15	100

Source: Data processing, 2022

Table 3. Differences in the area of land cover classes using the random forest method and the maximum likelihood classification

No	Land Cover	Random Forest (Ha)	Maximum Likelihood Classification (Ha)
1	Primary Dryland Forest	3886.61	4173.05
2	Secondary Dryland Forest	1644.61	1752.01
3	polder	303.65	241.99
4	Irrigation Field	415.99	326.66
5	Meadow	155.02	169.11
6	thicket	743.99	147.19
7	Field	126.76	310.27
8	Plantation	195.33	254.11
9	Built-up Land	368.45	458.23
10	Industrial Buildings	154.99	91.27
11	Mining	221.54	213.54
12	Gosong River	23.06	90.56
13	River	46.11	58.16
	Amount	8286.11	8286.15

Source: Data processing, 2022

IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- _____
Vol 4 No 1 | June 2023

3.1.3 Accuracy test

Table 4 Calculation of maker and user accuracy based on the random forest method confusion matrix

Land Cover	User Accuracy (%)	Error Commission (%)	Builder Accuracy (%)	Error Omission (%)
Primary Dryland Forest	100	0	60	40
Secondary Dryland Forest	77.77	22.23	53.84	46.16
polder	68.75	31.25	100	0
Irrigation Field	100	0	81.25	18.75
Meadow	100	0	100	0
thicket	71.42	28.58	90.90	9.1
Field	83.33	16.67	71.42	28.58
Plantation	80	20	100	0
Built-up Land	100	0	100	0
Industrial Buildings	100	0	100	0
Mining	100	0	100	0
Gosong River	100	0	100	0
River	100	0	100	0
Overall accuracy		86%		
Kappa		0.84		

Source: Data processing, 2022

Table 5 Calculation of maker and user accuracy based on the maximum likelihood classification method confusion matrix

Land Cover	User Accuracy (%)	Error Commission (%)	Builder Accuracy (%)	Error Omission (%)
Primary Dryland Forest	75	25	50	50
Secondary Dryland Forest	50	50	50	50
polder	38.46	61.54	100	0
Irrigation Field	60	40	76.44	23.56
Meadow	60	40	33.33	66.67
thicket	90	10	45	55
Field	50	50	100	0
Plantation	30.76	69.33	100	0
Built-up Land	80	20	26.66	73.34
Industrial Buildings	66.66	33.34	50	50
Mining	75	25	100	0
Gosong River	22.22	77.78	100	0
River	100	0	100	0
Overall accuracy		55%		
Kappa		0.51		

Source: Data processing, 2022

3.2 Research Discussion

3.2.1 Classification of land cover random forest method

Land cover mapping using Landsat 9 imagery data can be carried out using the random forest method with the random trees classifier train tool found in ArcGIS 10.8. Based on the results of processing, 13 classification classes of land cover were obtained, namely Primary Dryland Forest, Secondary Dryland Forest, Polder, Irrigated Paddy Fields, Grassland, Thickets, Fields, Plantations, Built-up Land, Industrial Buildings, Mining, River Basins, and Rivers. . The dominant land cover in the random forest classification is Primary Dryland Forest, covering 3886.61 Ha of the area or 47% of the total area in Lubuk Kilangan District. The land cover with the smallest area is Gosong Sungai, which is 23.06 Ha of the area or 0% of the area in Lubuk Kilangan District.

IRSAJ
International Remote Sensing Applied Journal
e_ISSN = 2775-3409
p_ISSN =_____- _____
Vol 4 No 1 | June 2023

3.2.2 Land cover classification using the maximum likelihood classification method

Land cover mapping using Landsat 9 imagery data was carried out using the maximum likelihood classification method with the maximum likelihood classification tools found in ArcGIS 10.8. There were 13 classification classes of land cover, namely Primary Dryland Forest, Secondary Dryland Forest, Polder, Irrigated Paddy Fields, Grassland, Thickets, Fields, Plantations, Built-up Land, Industrial Buildings, Mining, River Basins, and Rivers. The dominant land cover in the maximum likelihood classification is Primary Dryland Forest, covering 4173.05 Ha of the area or 50% of the total area in Lubuk Kilangan District. The land cover with the smallest area is the river, covering 58.16 ha of the area or 1% of the area in Lubuk Kilangan District.

3.2.3 Accuracy test

The accuracy test of the classification results was carried out to test the level of accuracy of the maps produced from the digital classification process with test samples from the results of field activities. The method used to calculate classification accuracy uses Kappa accuracy with the help of a confusion matrix. The confusion matrix is a comparison between the classified land cover in the image and the land cover in the field. Meanwhile, the Kappa value is not only determined by objects that are correctly classified but also take into account misclassification (Congalton & Green, 1999).

The results of the random forest land cover map accuracy test using the confusion matrix show an overall accuracy value of 86%, which means that 86% of the pixels are classified correctly. Meanwhile, the Kappa value obtained is 0.84, which this accuracy is included in the Almost Perfect Agreement category. The random forest classification method produces a land cover classification when viewed based on the standards used in this study, so the accuracy value is included in the feasible category. according to the accuracy suitability category according to Viera and Garrett (2005). Meanwhile, the results of the maximum likelihood classification method land cover map test show an overall accuracy value of 55%, which means that 55% of pixels are classified correctly. and, the Kappa value obtained is 0.51, which this accuracy is included in the Moderate Agreement category. The maximum likelihood classification method produces a land cover classification based on the standards used in this study, so the accuracy value is in the decent enough category which indicates that the maximum likelihood classification method is by the accuracy suitability category according to Viera and Garrett (2005).

4. CONCLUSION

a) Classification of land cover using the random forest method with 13 identified land cover classes found a total of 16,853 pixels and a total area of 8286.11 Ha with an overall accuracy value of 86% and a Kappa value obtained of 0.84. The training data used in the random forest method have been labeled with the appropriate land cover type, as well as spatial and spectral information from the observed objects. After training, the random forest model can be used to classify unlabeled remote sensing images with a fairly high degree of accuracy.
b) Land cover classification using the maximum likelihood classification method with 13 identified land cover classes found a total of 13,048 pixels and a total area of 8286.15 Ha with an overall accuracy value of 55% and a Kappa value obtained of 0.51. The training data used in the maximum likelihood classification method has been labeled with the appropriate land cover type, as well as spatial and spectral information from the observed object. After training, the maximum likelihood classification model can be used to classify unlabeled remote sensing images by determining the class that has the maximum probability from a predetermined class probability distribution.
c) Based on the comparison of the accuracy of the two methods being compared, it shows that the random forest method is better at classifying land cover than the maximum likelihood classification method. This is because the random forest method is very effective in dealing with overfitting. After all, the resulting trees or classifieds are done randomly. Meanwhile, the maximum likelihood classification method requires the maximum probability of the class probability distribution that has been determined. Random forest requires a long computational time to carry out the classification process. Meanwhile, the maximum likelihood classification method is simpler and faster in processing. However, it requires a condition where the way the data is spread in a population or sample fits the data.

5. REFERENCES

[1] Apriliadi, MA, & Kamaruddin, T. (2019). Classification of Land Cover Using Landsat 8 Imagery in Banda Aceh City and Aceh Besar District. Geosphere Education Journal, 4(2).
[2] Ayuindra, M. (2013). Land Cover Analysis Using Supervised And Unsupervised Classification
[3] Breiman L (2001). "Random Forests". Machine Learning. 45(1): 5–32. Doi:1023/A:1010933404324
[4] Burkov, Andriy. (2019). The Hundred-Page Machine Learning Book
[5] Danoedoro, P., & Murti, SH (2021). Classification of Land Cover/Use of Landsat-8 Oil Data Using the Random Forest Method. Indonesian Journal of Remote Sensing, 3(1), 1-7.
[6] Degrees, RM, Sopariah, Y., Aprilianti, S., Taruna, AC, Tisna, HAR, Ridwana, R., & Sugandi, D. (2020). Classification of Land Cover Using Landsat 8 Imagery Operational Land Imager (Oil) in Pangandaran District. Journal of Ocean Geography, 3(1), 1-10.
[7] El Naqa, I., & Murphy, MJ (2015). What Is Machine Learning?. In Machine Learning In Radiation Oncology (Pp. 3-11). Springer, Cham.
[8] Hamdir, ANRW (2014). Comparative Study of Maximum Likelihood Multispectral Classification and Support Vector Machine for Land Cover Mapping. Indonesian Earth Journal, 3(4).
[9] Ho, Tin Kam (1995). Random Decision Forests (Pdf). Proceedings Of The 3rd International Conference On Document Analysis And Recognition, Montreal, Qc, 14–16 August 1995. Pp. 278–282.
[10] Indonesia, SN (2010). Land Cover Classification. Jakarta. Indonesia.
[11] Iqbal, M. Landsat 8 Image Accuracy and Classification Test, Bima Regency, West Nusa Tenggara.
[12] Maksum, ZU, Prasetyo, Y., & Haniah, H. (2016). Comparison of Land Cover Classification Using Object-Based Classification Methods and Pixel-Based Classification in High and Medium Resolution Imagery. Undip Journal of Geodesy, 5(2), 97-107.
[13] Ralf Herbrich & Thore Graepel. 2018. Introduction To Machine Learning With Applications In Information Security. California: Crc Press.
[14] Richards, J. Remote Sensing Digital Image Analysis, Berlin: Springer-Verlag (1999), 240 Pp
[15] Sampurno, RM, & Thoriq, A. (2016). Land Cover Classification Using Landsat 8 Operational Land Imager (Oli) Imagery in Sumedang Regency (Land Cover Classification Using Landsat 8 Operational Land Imager (Oli) Data In Sumedang Regency). Journal of Teknotan Vol, 10(2).
[16] Septiani, R., Citra, IPA, & Nugraha, ASA (2019). Comparison of Supervised Classification and Unsupervised Classification Methods for Land Cover in Buleleng Regency. Journal of Geography: Information Media Development and the Geography Profession, 16(2), 90-96.
[17] Widiagani, A. (2017). Comparison of Decision Tree, Random Forest, and Rotation Forest Methods (Doctoral Dissertation, Indonesian University of Education)
[18] Yani, N., Sulistiawaty, S., & Usman, U. (2022). Comparison of Maximum Likelihood Classification Methods and Support Vector Machine Learning in Flood Inundation Mapping (Study Area: Lake Tempe Region). Indonesian Journal of fundamental sciences, 8(1), 9-17.