JTKSI, Vol.04 No.03 September 2021 ISSN : 2620-3022
Hal. 130-134

Paddy Growth Stages Classification based on Hyperspectral Image using Decision Tree and Naive Bayes

Febri Maspiyanti[1], Ghinanti Pratiwi[2], Jullend Gatc[3]
[1,2]Informatics Engineering, Universitas Pancasila, Indonesia
[3]Information System, Kalbis Institute, Indonesia
[1,2]Jl. Raya Lenteng Agung No.56-80, Srengseng Sawah, Jagakarsa, Jakarta, Indonesia.
[3]Jl. Pulomas Selatan kav.22, Jakarta Timur, Jakarta, Indonesia
Email: febri.maspiyanti@univpancasila.ac.id

Received: 28 Juli 2021 Revised: 17 Agustus 2021 Accepted: 19 Agustus 2021

Abstract-Hyperspectral imaging is one of remote sensing technology that gather information from a wide spectrum of electromagnets called spectral bands, with the aim of finding objects, identifying materials, or detecting processes. In an effort to calculate the amount of rice crops can be harvested within a certain periode of time, we need to accurately predict the growing phase of paddy plant at that time. In determining the phase of the rice plant with high accuracy value, need to be supported with the selection of appropriate algorithms, and also the features selection. In this study, a comparison between the Decision Tree and Naive Bayes methods to classify the nine phases of rice growth based on hyperspectral image achieve accuracy value of 91.67% and 83% respectively. Based on the accuracy result, our new proposed method improved 6,38% accuracy compare to our previous research.

Keyword: Classification, Decision Tree, Growth Stage, Hyperspectral, Naïve Bayes

I. INTRODUCTION

Indonesia is one of the largest agricultural rice producing countries in the world. Based on Badan Pusat Statistik (BPS-Statistics Indonesia) in 2017, Indonesia produces 81,073 tons of rice [1]. For most Indonesian people, rice is a primary food crop that is very important, especially in daily life. Time series process and accurate extraction of rice distribution can bring vital information for national food security, agricultural policy formulation, and regional environmental sustainability [2]. This condition will also affect other sectors because the food sector must be sustainable enough to sustain the welfare of other sectors.

In the process of data collecting, field officer of the district's agricultural bureau usually estimates the harvest area by collect data directly at some rice field sample areas. The collecting and compiling data of the district to national level using conventional method is time consuming activity. In addition, the data collected up to the national level lacks its validity because the estimated harvested land in the district is neither accurate nor meticulous [3].

The development of remote sensing technology along with the hyperspectral image is able to overcome the obstacle in determining of paddy growth stages [4]. One of the utilizations of remote sensing technology along with hyperspectral image in paddy farming is in the development of harvesting estimation models by determining the growth phase of paddy in certain period. Hyperspectral imaging is able to distinguish the growing phase of plants based on its reflectance spectrum [5] [6]. Therefore, the layers of hyperspectral image are the features that contain reflectance values that determine the growth stage of paddy. Paddy growth stages based on the International Rice Research Institute (IRRI) divided into nine phases, namely: Vegetative 1, Vegetative 2, Vegetative 3, Reproductive 1, Reproductive 2, Reproductive 3, Ripening 1, Ripening 2, and Ripening 3 [4] [7] [8] [11]. In the aim of determine paddy growth stage with optimal accuracy value, it need to be supported not only by the fitted algorithm but also by the right features [8]. The aim of this study is to classify extracted hyperspectral image into nine phases paddy growth stages based on IRRI using Decision Tree and Naïve Bayes.

II. RESEARCH METHOD

Hyperspectral sensors provide more imagery alternatives, and newly developed image processing algorithms provide more analytical tools, hyperspectral remote sensing is positioned to become one of the core technologies for geospatial research, exploration, and monitoring

[9]. Ç. Küçük [10] build a feasible phenology classification schema for paddy-rice using multitemporal co-polar TerraSAR-X images. Phenology 6 classes classification was conducted with support vector machines (SVM), k-nearest neighbors (kNN), and decision trees (DT). Küçük research obtain above 80% of accuracy. Takayama [11] propose semi-supervised classification method with hyperspectral data using SVM and Sparse Linear Discriminant Analysis (SDA) in 6 classes paddy growth stages data. Takayama's research achieve up to 89,3% of accuracy. Singha [12] using MODIS NDVI data to do two classes (rice and non-rice) classification achieve up to 93% of accuracy. Halim [13] use PCA and Kernel PCA to reduce the dimension of 6 classes hyperspectral data. Halim research achieve up to 93,33% of accuracy.

A. Data

The data used in this study is the hyperpectral image of the rice field of Subdistrict of Indramayu West Java Indonesia, which is also the data from previous research [4] [8] [14]. We use 1800 dataset consists of 9 classes with 126 bands that have been filtered of 10 bands due to water absorption in the atmosphere [3] leaving 116 bands remaining. The remaining 116 bands data, knowledge-based feature selection was performed in accordance with Maspiyanti et al [4] hence remaining the 4 bands data.

Our 1800 data consist of balanced classes (veg1=200 data, veg2=200 data, veg3=200 data, rep1=200 data, rep2=200 data, rep3=200 data, rip1=200 data, rip2=200 data, rip3=200 data). The raw data can be seen in fig. 1, while the extracted data can be seen in fig. 3. The extracted data are the reflectance value between 0 to 0.55. In fig. 2 we can see the graphic developed by one of the pixels from raw data (fig. 1) using ENVI [15].

Figure 1. Indramayu hyperspectral imaging data

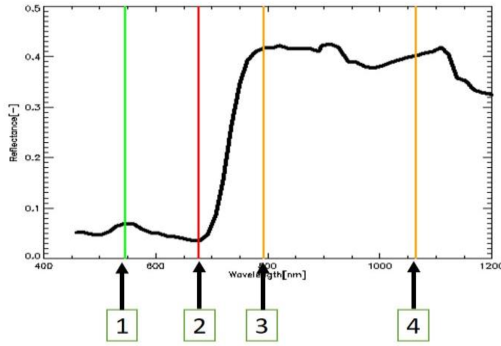

Figure 2. Z profile from hyperspectral imaging

As seen in fig. 2, the x axis are the wavelenghts in nanometer, while the y axis are the reflectances value for each x wavelength. From fig. 2, the green line (1) indicate the visible green light, the red line (2) indicate the visible red light, the first orange line (3) indicate the start of near infrared (NIR), and the second orange line (4) indicate the end of NIR. The cpmbination of visible lights (1 and 2) indicate the leaf colour, while the NIR (3 and 4) indicate the cell structure condition inside the plant (include the photosynthesis process). In this study, we focused on these 4 wavelengths value in order to distinguish one class to other classes.

The high value of NIR (higher than 0.4) indicate the high photosynthesis process, while the lower NIR value indicate the early growing phase or the late of growing phase (nearly harvest phase) which has medium or low photosynthesis process because of the lack of leaf number (early growing phase) or the amount of chlorophyll is decreased due to leaf aging (nearly harvest time).

0.0462	0.0419	0.0419	0.0397	0.0468	0.0554	0.0609	0.0607	0.0529	0.0476	0.0448
0.0476	0.0455	0.0423	0.0412	0.0456	0.0551	0.06	0.0587	0.0511	0.0465	0.0435
0.0476	0.0436	0.043	0.0407	0.0461	0.0556	0.0593	0.0594	0.0518	0.0465	0.0445
0.0546	0.0494	0.0482	0.0473	0.0544	0.068	0.0741	0.0747	0.0678	0.063	0.0615
0.0407	0.0426	0.0417	0.0391	0.0436	0.0565	0.0616	0.0601	0.0522	0.0462	0.0438
0.0441	0.0436	0.0397	0.0401	0.0451	0.0543	0.0597	0.0583	0.0518	0.045	0.0431
0.048	0.0448	0.0419	0.0418	0.0459	0.0563	0.0617	0.0603	0.0533	0.0481	0.0455
0.0459	0.0417	0.0417	0.0417	0.0456	0.0543	0.0593	0.0583	0.0511	0.0457	0.0445
0.0476	0.0436	0.0423	0.0401	0.0456	0.0556	0.0608	0.0594	0.0522	0.0469	0.0442
0.048	0.0439	0.0419	0.0413	0.0449	0.0554	0.0609	0.0592	0.0522	0.0462	0.0445
0.0427	0.0419	0.0413	0.0397	0.0434	0.0541	0.0589	0.0596	0.0514	0.0462	0.0434
0.0427	0.0419	0.0419	0.0418	0.0459	0.0549	0.0605	0.0603	0.0525	0.0473	0.0455
0.0459	0.0417	0.0403	0.0396	0.0436	0.0551	0.0604	0.0594	0.0522	0.0462	0.0445
0.0452	0.047	0.0424	0.0413	0.0457	0.057	0.0616	0.0596	0.052	0.0462	0.043
0.0425	0.0436	0.0423	0.0407	0.0446	0.0551	0.0593	0.0583	0.0518	0.0457	0.0435
0.0494	0.0455	0.0423	0.0417	0.0456	0.0551	0.0593	0.0587	0.0518	0.0462	0.0442
0.048	0.0439	0.0419	0.0418	0.0468	0.0549	0.0605	0.0592	0.0518	0.0454	0.0437
0.0459	0.0436	0.0397	0.0396	0.0441	0.0561	0.0604	0.0587	0.0514	0.0453	0.0431
0.048	0.0439	0.0399	0.0381	0.0439	0.0549	0.0605	0.0584	0.0503	0.045	0.0416

Figure 3. Extracted data in .txt format file

B. Method

In this study, we use Decision Tree (DT) C 4.5 and Naïve Bayes (NB) classifier to classify 9 classes hyperspectral paddy data. For the DT classifier, we combine our previous study by using the range of feature value from Maspiyanti [4]. Our proposed method can be seen in fig. 4 below:

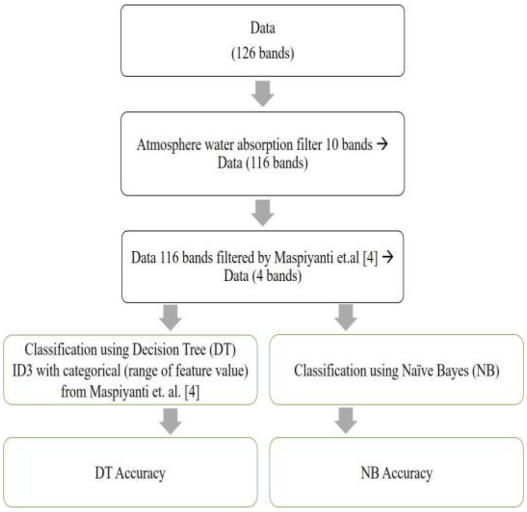

Figure 4. Proposed Research Method

The pre-processing method has been done in our previous study [4] [8] [14], hence this study focused on the classification improvement. For the DT classifier, we use DT [16] [17] with entropy and Gain Information formula as described below:

In order to develop DT model, in the first place we need to specify the Root Node, Child Nodes, and the Leaves Node. To specify the Selected Node (current node), compute the entropy value of each feature using training data, then compute the Information Gain value. The feature with the largest Information Gain value will become the selected node.in order to find Entropy value, use formula given:

$$Entropy(S) = -p_+ \log_2 p_+ - p_- \log_2 p_- \quad (1)$$

where:
S : Certain class
(p+) : probability of positive class
(p-) : probability of negative class

Once we obtain the Entropy value, compute the Information Gain, as given formula:

$$\text{Gain}\ (S, A) = \text{Entropy}\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(Si) \quad (2)$$

where:
S : Certain class
A : Certain Feature
n : number of A
|Si| : number of i^th of S
|S| : total number of S

Naive Bayes, for each decision class, calculate the probability on condition that the decision class is correct, according to the object information vector given. This algorithm assumes that the object attribute (feature) is independent. Naïve Bayes algorithm described below:

1. The training process: compute *mean* and *standard deviation:*

a. Compute *mean* for each feature and class, as given formula:

$$\mu = \frac{\sum x_i}{n} \quad (3)$$

where:
μ : mean
n : number of data
$\sum x_i$: sum of data

b. Compute *standard deviation* for each feature and class, as given formula:

$$\sigma = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2\right)^2 \quad (4)$$

where:
σ : standard deviation
μ : mean
n : number of data
$\sum x_i$: sum of data

2. Testing Process:

a. Find probability density from the data testing using given normal distribution formula:

$$\text{P}\ (X_i = x_i | Y_i = y_i) = \frac{1}{\sqrt{2\pi\sigma ij}} \text{e}^{-\frac{(x_i - u_{ij})2}{2\sigma^2 ij}} \quad (5)$$

where:
P : probability
Xi : i^th feature
xi : i^th feature value
Yi : Certain class
yi : certain sub-class
π : *phi*
μ : *mean* of the entire feature
σ : *variance* of the entire feature

b. Next step is to find Likelihood function, as given formula:

$$P(X_i | Y) = \prod_{i=1}^{n} P(X_i | Y) \quad (6)$$

where:
P(X|Y): Probability distribution of each X given Y
X_i : i^th attribute or feature
Y : Certain class

Once we have the Likelihood value, we can compute the probability of each class in order to find the final probability. The NB classification process can be seen in fig.5.

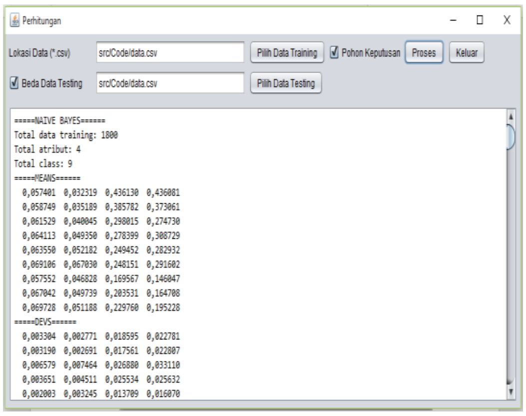

Figure 5. Classification process of 4 features using NB

C. Evaluation

In this study, we use accuracy as the performance measurements method. The accuracy formula can be seen below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

where TP is True Positive (positive classified as positive), TN is True Negative (negative classified as negative), FP is False Positive (negative classified as positive), FN is False Negative (positive classified as negative) [19].

III. RESULTS

Our proposed method result can be seen in table 1 below:

Table 1. Accuracy

Classifier	Classification		
	Correct	***Incorrect***	***Accuracy (%)***
DT	1650	150	91,67
NB	1508	292	83

In table 1, can be seen the accuracy of DT achieve highest value of 91,67%. From 1800 data, 1650 data correctly classified using DT, and 1508 data correctly classified using NB. Hence, in this study, DT achieved highest number of accuracies. From all previous study that use the same Indramayu hyperspectral image data, this study shows the highest accuracy. We combine DT with range of feature value knowledge based on Maspiyanti et. al. [4]. We use Maspiyanti et. al. [4] fuzzy logic membership as the range of feature value for each data. Hence, the accuracy improved by 6,38%.

IV. CONCLUSION

By using DT classifier combined by knowledge based provided by our previous study [4], this study achieve accuracy of 91,67% which vanquishes other previous studies that using the same Indramayu hyperspectral image data. As mentioned before, that in the aim of determine paddy growth stage with optimal accuracy value, it needs to be supported not only by the fitted algorithm but also by the right features.

REFERENCES

[1] BPS-Statistics Indonesia and Directorat General of Food Crops, "Produksi, luas panen dan produktivitas Padi di Indonesia 2013-2017," Ministry of Agriculture Republic of Indonesia, Jakarta, 2017.

[2] Xu, Xinjie, et. al. "Evaluation of One-Class Support Vector Classification for Mapping the Paddy Rice Planting Area in Jiangsu Province of China from Landsat 8 OLI Imagery," Remote Sensing, vol. 10, no. 4, p. 546, 2018.

[3] Mulyono, Sidik, M.I. Fanany, and T. Basaruddin. A paddy growth stages classification using MODIS remote sensing images with balanced branches support vector machines. ICACSIS. IEEE. 2012.

[4] F. Maspiyanti, M. I. Fanany and A. M. Arymurthy, "Paddy Growth Stages Classification based on Hyperspectral Image using Modified Fuzzy Logic," Jurnal Penginderaan Jauh dan Pengolahan Data Citra Digital, vol. 10, no. 1, pp. 41-48, 2013.

[5] N. Kosaka, S. Miyazaki, U. Inoue, "Vegetable green coverage estimation from an airborne hyperspectral image", Geoscience and Remote Sensing Symposium 2002. IGARSS '02. 2002 IEEE International, vol. 4, pp. 1959-1961 vol.4, 2002.

[6] El-Hendawy S, Al-Suhaibani N, Hassan W, Tahir M, Schmidhalter U. Hyperspectral reflectance sensing to assess the growth and photosynthetic properties of wheat cultivars exposed to different irrigation rates in an irrigated arid region. PLOS ONE 12(8): e0183262. 2017.

[7] International Rice Research Institute (IRRI). Paddy Growth Stages 0-9 Phase <http://www.knowledgebank.irri.org/extention/growth-stages-0-9.html>.

[8] Paddy Growth Stages Classification Based on Hyperspectral Image using Feature Selection Approach. The 14th International Conference on Quality in Research. ISSN:1411-1284. 2015.

[9] Sippert, Peg. Why Use Hyperspectral Imagery. Journal of Photogrammetric Engineering & Remote Sensing. Pp. 377-380. 2004.

[10] Ç. Küçük, G. Taşkın and E. Erten, "Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 6, pp. 2509-2519. IEEE. 2016.

[11] T. Takayama, N. Yokoya and A. Iwasaki, "Optimal hyperspectral classification for paddy field with semisupervised self-learning," 2015 7th Workshop on Hyperspectral Image and Signal

Processing: Evolution in Remote Sensing (WHISPERS). IEEE. 2015.
[12] Singha, Mrinal, Bingfang Wu, and Miao Zhang. An Object-Based Paddy Rice Classification Using Multi-Spectral Data and Crop Phenology in Assam, Northeast India. Remote Sensing. MDPI. 2016.
[13] Halim, H., S. M. Isa and S. Mulyono, "Comparative analysis of PCA and KPCA on paddy growth stages classification," 2016 IEEE Region 10 Symposium (TENSYMP). IEEE. 2016.
[14] Suhandono, Nugroho, Febri Maspiyanti, M. I. Fanany. Extreme Learning Machine for Growth Stages Classification of Rice Plants from Hyperspectral Images Subdistrict Indramayu. KCIC. 2013.
[15] Harris Geospatial Solutions: ENVI. http://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx.
[16] K. Jearanaitanakij, "Classifying Continuous Data Set by ID3 Algorithm," 2005 5th International Conference on Information Communications & Signal Processing. IEEE. 2005.
[17] Bishop, Christopher. Pattern Recognition and Machine Learning. Springer-Verlag New York. 2006.
[18] D.L. Olson, D. Delen, Advanced data mining techniques, Springer Publishing Company Inc. 2008.
[19] Baratloo, Alireza, et. al. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. Emergency. NCBI. 2015.