Article

Using satellite image data to identify rice varieties through linear spectral unmixing method (case study: Karangjati Sub District, Ngawi Regency)

Moch Rafli Kusoiry¹, Lalu Muhamad Jaelani¹,*, Hartanto Sanjaya²,³

¹Department of Geomatics Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
²Department of Civil Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
³National Research and Innovation Agency, Jakarta Pusat 10340, Indonesia

* Corresponding author: Lalu Muhamad Jaelani, lmjaelani@its.ac.id

Abstract: Remote sensing technology has increasingly emerged as a potent tool for precision agriculture, particularly in facilitating the mapping and monitoring of crops on a large scale. An application of this technology is the identification of different types of rice by analyzing the pixels acquired in satellite images. Regrettably, the pixels in the image have been mixed from different recorded items. Therefore, they have the potential to influence the outcome of the identification. An effective approach to address this problem is to employ the linear spectral unmixing (LSU) technique. The LSU approach quantifies the ratio of pure objects in every pixel of an image by utilizing the spectral value associated with the endmember of the rice variety. The investigation was carried out in the Karangjati District during the generative stage (70 ± DAP) of the rice planting season. The data indicates that the dominant variety is Inpari 32 HDB. The data validation tests, which involved the use of a confusion matrix and Kappa analysis, resulted in an overall accuracy rate of 85.48% and a Kappa analysis score of 70.6%.

Keywords: endmember; Karangjati; linear spectral unmixing; precision agriculture; remote sensing; rice varieties

1. Introduction

Since rice, or the grain of the rice plant, is the primary meal of the Indonesian people, rice is a vital food crop commodity in Indonesia. Indonesian rice has distinctive characteristics depending on the type of variety [1]. The rice-growing process involves multiple phase: 1) the planting phase, or it can be called the initial phase of growth, is when waterlogging activities in paddy fields are carried out; 2) vegetative phase: in this phase, the paddy fields are dominated by green color due to leaf growth; 3) the generative phase is when paddy fields have started to rinse rice grains, and the leaves start to turn yellow; thus, the paddy fields begin to turn yellow [2]. In Indonesia, almost 95% of people consume rice as a staple food. Also, rice is the primary commodity in supporting people’s food [3]; thus, every year, the demand for rice increases in line with the increase in population [4]. Since Indonesian people’s rice consumption level is significant, it must also be balanced with balanced production efforts to avoid the risk of a shortage of supplies, which impacts food insecurity [5]. One of the efforts of the Indonesian state in dealing with this is by controlling rice production based on its variety.

Remote sensing is a useful tool for investigating different types of rice. In order to perform research on rice varieties, we need imagery from satellites that have high spatial, spectral, and temporal resolution. We opted to utilize Sentinel-2 and MODIS
imagery due to its favorable specifications in terms of temporal and spectral resolution [6–12].

However, some obstacles arise when choosing remote sensing satellite imagery as a medium for conducting this research. One pixel on a multispectral remote sensing satellite image usually contains more than one type of recorded feature, resulting in mixed pixels. These mixed pixels affect the accuracy of identification. The spectral unmixing method can be used to deal with mixed pixels. To deal with the mixed pixel, there are two general approaches used by researchers: a linear spectral unmixing (LSU) that assume the photon only interact with one substance on the surface, and nonlinear spectral unmixing (NSU) that assume the photon has multiple reflection or interact with a variety of substances. In fact, the NSU is very close to reality, but the LSU implementation is more simple with accurate results when dealing with moderate spatial resolutions [13,14]. LSU is a method for identifying the percentage of the presence of a pure object in each image pixel based on its spectral value, which is called an endmember. In this case, the intended endmember is rice varieties. The LSU method is also applied to multispectral images [6,15,16]. Based on the background above, this research was conducted using LSU.

The purpose of this study was to find out what varieties of rice are planted in the Karangjati sub-district, which varieties are dominant, and to find out the accuracy of this study using Sentinel-2 imagery data with the linear spectral unmixing method.

2. Material and methods

2.1. Study area

This research was conducted in Karangjati Sub District in Ngawi Regency, which included vector data of paddy fields. Karangjati Sub District has geographical coordinates of 7°27′31″ South Latitude and 111°36′31″ East Longitude, as shown in Figure 1.

![Figure 1. The location of study area.](image-url)
2.2. Data

The data utilized in this investigation are categorized into three distinct groups of data. The primary source of data is satellite imagery, specifically the MODIS MCD43C4 dataset, which spans the whole research region from January 2022 to March 2023. Additionally, the Harmonized Sentinel-2 MSI: Multispectral Instrument Level-2A dataset acquired on 13 January 2023, covering the entirety of the research area. The growth stage of rice can be determined using MODIS MCD43C4 satellite imagery. The satellite’s high temporal resolution of 1–2 days makes it a dependable tool for monitoring the rice growing phase. Three endmembers that represent different rice varieties/types at the field (i.e., Ciherang, Cibogo, and Inpari 32 HDB) were recorded on 3–4 February 2023, following the work protocol of Sanjaya [17]. Meanwhile, the LSU approach utilizes Sentinel-2 satellite imagery to discern different types of rice by analyzing the endmember of each variety. This is made possible due to the high spatial and spectral resolution of the satellite imaging, as stated on Sentinel’s official website (sentinels.copernicus.eu). The Sentinel is equipped with a multispectral instrument that has 13 spectral bands spanning the visible, near-infrared, and short-wave infrared channels. The satellite possesses a spatial resolution of 10 meters for the visible and near-infrared bands, while for the near-infrared and short-infrared bands, the resolution is 20 m and 60 m respectively.

The acquisition of all satellite imagery was facilitated by the Google Earth Engine platform (earthengine.google.com). The GEE platform offers numerous advantages in terms of data processing, both in terms of hardware and software [18]. The second component consists of the endmember data acquired from the spectral recordings of each leaf of the rice variety cultivated in the study area. The Ocean Optics USB4000 spectrometer was used to conduct spectral recordings. The third dataset consists of vector data representing the unprocessed paddy fields in Karangjati Sub District, Ngawi Regency. The data is sourced from the Indonesian Geospatial Information Agency (BIG). Agricultural fields used for cultivating rice.

2.3. Workflow method

The stages of processing research data are shown in Figure 2. There are four important stages: determining the endmember of each rice variety, determining the threshold value of the generative phase of rice using MODIS, applying the threshold value to Sentinel-2, and inputting the endmember to Sentinel-2 to perform LSU.
2.4. Determine the endmember

To identify the endmembers of each rice variety, the spectral values of leaf samples are recorded using a spectrometer. After processing, the recorded spectral value results will be identified as an endmember for every rice variety during a particular rice growth phase. The endmember results will subsequently be applied to the generative phase of the Sentinel-2 band as an input of the LSU formula, since the sample was collected during the generative phase. To identify the endmembers of each rice variety, the spectral values of leaf samples are recorded using a spectrometer. After processing, the recorded spectral value results will be identified as an endmember for every rice variety during a particular rice growth phase. The endmember results will subsequently be applied to the generative phase of the Sentinel-2 band as an input of the LSU formula, since the sample was collected during the generative phase.

2.5. Determine the threshold value

Two spectral indices from MODIS image data, Normalized Difference
Vegetation Index (NDVI) and Normalized Different Water Index (NDWI) were used to establish the threshold value for identifying the generative phase from paddy fields. NDVI calculations are based on the principle that growing green plants can very effectively absorb radiation in the visible light spectrum (Photosynthetically Active Radiation or PAR), while green plants are highly reflective of radiation from the near-infrared region (NIR). The usefulness of the NDVI index in determining the value of this threshold is that in the generative phase of rice, the process of filling and ripening of the rice panicles occurs, indicated by the condition of the rice leaves starting to turn yellow and the appearance of yellow rice panicles thus that the fields begin to be dominated by yellow color. Based on this knowledge, the NDVI index can be used to determine the level of greenery in paddy fields. The lower the NDVI value, it can be concluded that the rice fields are starting to turn yellow. The concept of the spectral pattern is based on this principle using only the red band image as follows [19,20]:

$$\text{NDVI} = \frac{(\rho_{\text{NIR}} - \rho_{\text{Red}})}{(\rho_{\text{NIR}} + \rho_{\text{Red}})}$$

(1)

In contrast, water bodies can absorb strongly at visible and infrared wavelengths. An NDWI value greater than zero indicates that the surface contains water/water bodies, but conversely, if the value is smaller or equal to zero, it is assumed to be a non-water surface [21]. The use of NDWI in determining this threshold value follows the situation when in the field that the generative phase of rice is indicated by the elongation of rice leaves and the appearance of rice panicles so that leaves cover the condition of the rice fields along with the rice panicles, with the rice fields being covered causing part of the paddy soil that contains water to be covered. Then, the NDWI index can be used to determine the level of water presence in paddy fields, with the assumption that the lower the NDWI value, it can be concluded that the level of water presence in paddy fields begins to disappear. NDWI has an equation pattern as follows:

$$\text{NDWI} = \frac{(\rho_{\text{Green}} - \rho_{\text{NIR}})}{(\rho_{\text{Green}} + \rho_{\text{NIR}})}$$

(2)

2.6. Application of threshold values into Sentinel-2

At this stage, the Sentinel-2 data is updated with the generative phase threshold value. The “generative phase Sentinel-2 image data” that is produced as a result of this step only shows rice fields that are in the generative phase.

2.7. Input endmember into Sentinel-2 to perform LSU

The endmember is then input into Sentinel-2 to perform LSU, at this stage each rice variety endmember is input into the Sentinel-2 image data. In this case, the endmember targeted is the rice variety. Before inputting the endmembers, it is necessary to adjust the wavelength of each endmember to the Sentinel-2 image band. Then the LSU process is carried out. The results of this method are image data of endmember fractions from each rice variety studied. LSU is a method for identifying the percentage of pure objects present in each image pixel based on the spectral value which is called the endmember [22,23]. The standard technique for performing
spectral unmixing analysis is linear spectral unmixing (LSU) [24]. The LSU method is also applied to multispectral images [15].

2.8. Classifying the endmember fraction

Classifying the endmember fraction results of each variety: At this stage, the results of the endmember fractions are classified into four classes based on the range of minimum and maximum values of the fractions. The four classes are 0–25%, 25%–50%, 50%–75%, and 75%–100%. Each class indicates the level of existence of fractions of the endmember in each pixel. The larger the percentage value, the larger the presence of endmember fractions in a pixel (dominant), and the smaller the percentage value, the smaller the presence of endmember fractions in a pixel (not dominant).

2.9. Accuracy test

The accuracy of the processing results must be tested in this study. The percentage of accuracy can be validated by using accuracy tests to identify errors. A confusion matrix is utilized to conduct the accuracy test of a model’s or classification algorithm’s performance [25]. Within a matrix, the Confusion Matrix indicates how many of the model’s predictions were right and wrong. It is a square matrix that summarizes the result into the true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Accuracy test assessment can use a contingency matrix, which is a square matrix that contains the number of pixels that are classified. It has been determined that the lowest level of classification or interpretation accuracy using remote sensing is less than 85% [26].

3. Result and discussions

3.1. Endmember of rice varieties

Each variety of rice is measured for its spectral value during this process. Only the rice varieties grown at the study site were measured. A spectrometer was used to measure the spectral value of each leaf in each rice variety; approximately 90 leaves were measured for each variety. The midpoint of the leaf tip is the portion of the leaf body that is measured, as illustrated in Figure 3. The area of the leaf to be measured is indicated in the figure by the red circle around it. calculated with a spectrometer.

Figure 3. The illustration of rice leaf part for spectra measurement.

The Cibogo, Inpari 32 HDB, and Ciherang varieties provided the successfully measured spectral data. Following the measurement of each sample’s spectral values in the field, the results are filtered in an effort to remove data anomalies. Figure 4 shows an example of a rice variety data set that has been filtered for anomalies.
Advances in Modern Agriculture 2024, 5(2), 2538.

Figure 4. Measured spectra of rice leaf.

The data free of outliers for each rice variety is then averaged to produce pure endmembers based on the outcomes of the outlier filtering. The ideal endmember data is obtained after the average process is completed. A description of the data that was transformed into endmember data for every variety of rice examined is presented in Figure 5.

Figure 5. Endmember of each variety.

3.2. Identification of the rice generative phase from MODIS

In this study, it is necessary to guarantee that all data measured and processed are in the same phase, which is why the rice planting phase in the generative phase was identified. In terms of image data processing, it must be made sure that the data processed is limited to areas that are in the generative phase, as previously mentioned, since the field data measured is rice that is in this phase of growth. Take into account that the spectral values and endmember data of each object are unique. Different spectral values and endmember data will be present in rice during its
vegetative and generative phases, for instance. Using threshold values from NDVI and NDWI indices calculated from MODIS satellite imagery data based on 10 sample points at research locations over a one-year period, from January 2022 to March 2023, is the process employed in this procedure. Figures 6 and 7 display the NDVI and NDWI values, respectively, that were obtained. A value of 0 to 9 in Figures 6 and 7 are the code of samples.

![NDVI Value of MODIS](image1.png)

**Figure 6.** NDVI Value of MODIS in Karangjati Sub District.

![NDWI Value of MODIS](image2.png)

**Figure 7.** NDWI Value of MODIS in Karangjati Sub District.

To make the process of analyzing the threshold value easier, the two indices (NDVI and NDWI) are combined by averaging their respective index values. Figure 8 shows the combination’s outcomes. The average NDVI index value is displayed in blue, and the NDWI is displayed in orange.

From the combined NDVI and NDWI indices, the threshold value for the generative phase of rice plants was determined. The threshold value is determined by looking at the starting points of the NDVI and NDWI decline patterns, namely in June 2022 and October 2022. The results of the analysis show that the threshold values for the generative phase are NDVI > 0.7 and NDWI < −0.65.
3.3. Generative phase of rice in Sentinel-2A

The date of the Sentinel-2 image data used is 13 January 2023. The selection of dates is dependent upon factors related to cloud cover, and the image data must be in the generative phase at the time of each rice leaf sample’s spectral value measurement. Figure 9 displays the results of the first step, which involved substituting the Sentinel-2A image data with the rice field vector data in the Karangjati Sub District.

The generative phase threshold value is then used to carry out the image data masking procedure. Figure 10 shows the results of the masking process, which yields image data with rice areas in the generative phase. The reddish area is not in the generative phase, whereas the green area shows that the region is. It is important to keep in mind that not every paddy field in the Karangjati Sub District is currently growing.

Figure 9. Sentinel-2A imagery data subsetted by the Raw Paddy Field Vector Data.
Figure 10. Generative phase masking results.

3.4. Endmember fraction of rice varieties

The generative phase masking process is then carried out by inputting the endmembers of every variety of rice into the linear spectral unmixing (LSU) method. A range of distribution values for each endmember in the image data is the outcome of this process. The proportion of the endmember object is dominant if the range of values is near to 1. On the other hand, if the value range is near zero, it indicates that the object’s endmember proportion is not dominant. Figure 11 shows the distribution of all final member fractions for the Ciherang, Cibogo, and Inpari 32 HDB varieties. The dominant varieties (composite from three varieties used in this study) were indicated by green color, the less dominant one by yellow color.

Figure 11. Endmember distribution of three varieties composite.
The results of the range of endmember fraction values for each rice variety are shown in Table 1.

Table 1. Range of endmember fraction values for each rice variety.

<table>
<thead>
<tr>
<th>Varieties</th>
<th>Endmember Fraction Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
</tr>
<tr>
<td>Inpari 32 HDB</td>
<td>0.510</td>
</tr>
<tr>
<td>Ciperang</td>
<td>0.000</td>
</tr>
<tr>
<td>Cibogo</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Based on Table 1’s endmember value range, data indicates that the Inpari 32 HDB variety has a value range of 0.510 to 0.941, indicating a high occurrence rate of this variation in certain pixels. There are Ciperang and Cibogo varieties with low occurrence rates in some pixels because their values are below 0.5, despite having nearly the same range of values, which is between 0 and 0.25. Analysis is done using Table 1’s results. The Inpari 32 HDB variety is the most prevalent rice variety at each pixel, according to the analysis’s findings based on the statistical data in Table 1, whereas the Ciperang and Cibogo varieties are the least prevalent. Thus, it can be said that the Inpari 32 HDB variety is the most widely used variety in Karangjati Sub-District.

The dominant endmember range values are then used to group rice varieties. To distinguish the dominant HDB Inpari 32 variety from non-dominant varieties, this was done. Ciperang, Cibogo, and other varieties are the non-dominant types in question. Given that the “MR” variety was discovered to have been planted in the Karangjati Sub District based on the findings of the field survey, it is possible that other varieties exist as well. The dominant endmember range values are then used to group rice varieties. To distinguish the dominant HDB Inpari 32 variety from non-dominant varieties, this was done. Ciperang, Cibogo, and other varieties are the non-dominant types in question. Given that the MR variety was discovered to have been planted in the Karangjati Sub District based on the findings of the field survey, it is possible that other varieties exist as well.

The threshold value used as a reference to obtain information regarding the presence of the dominant HDB Inpari 32 variety in the fraction image is more than 0.50 or > 0.5. Meanwhile, the threshold value used to detect the presence of non-dominant varieties in the fraction image is less than 0.50/<0.5. The principle of using threshold values using these range values is that if the endmember value of a rice variety in one pixel is above 0.50, then that pixel is confirmed to contain the Inpari 32 HDB variety, but if the value is below 0.50, then that pixel contains varieties other than Inpari 32 HDB (Ciperang, Cibogo, and other varieties). The results of these reference values can be seen in Figure 12.
The endmember fraction classification procedure is then implemented according to the range value of every variety of rice. The classification is grouped into four classes: 0–25%, 25%–50%, 50%–75%, and 75%–100%. Each class indicates the level of existence of the final member shard in each pixel. The larger the percentage value, the greater the presence of endmember fractions in a pixel (dominant); the smaller the percentage value, the smaller the presence of endmember fractions in a pixel (not dominant). The results of the classification of the percentage of endmembers in Inpari 32 HDB, Ciherang, and Cibogo varieties can be seen in Figures 13–15.

**Figure 12.** Result of endmember fraction of each variety.

**Figure 13.** Classified endmember fraction of Inpari 32 HDB variety.
The results of calculating the area based on GEE statistical data for each variety are listed in Table 2. Based on Table 2, the area containing the Inpari 32 HDB variety is 1095 hectares. Meanwhile, the area of rice fields planted with other varieties, namely Ciherang, Cibogo and others, is 1184 hectares.

Table 2. The area of the pixels containing the endmember of each variety.

<table>
<thead>
<tr>
<th>Varieties</th>
<th>Area (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpari 32 HDB</td>
<td>1095</td>
</tr>
<tr>
<td>Ciherang, Cibogo, etc.</td>
<td>1184</td>
</tr>
</tbody>
</table>
3.5. Accuracy test

Following completion of all data processing results using the LSU method, an accuracy test utilizing the confusion matrix method takes place. This test includes Kappa analysis, producer accuracy, user accuracy, and overall accuracy values. For the purpose of using the sample point coordinates for accuracy testing, researchers conducted a field survey for each variety of rice. They were able to obtain 62 sample point locations for each variety. In Figure 16, the distribution of all sample points is shown.

![Figure 16. Distribution of validation sample points.](image)

Then the 62 sample points began to be calculated for the accuracy test using the confusion matrix, the accuracy test for the confusion matrix is shown in Table 3.

**Table 3. Confusion matrix table.**

<table>
<thead>
<tr>
<th>Procedure’s accuracy/commission error</th>
<th>Class</th>
<th>Inpari 32 HDB</th>
<th>Other varieties</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpari 32 HDB</td>
<td>30</td>
<td>4</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Other varieties</td>
<td>5</td>
<td>23</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>35</strong></td>
<td><strong>27</strong></td>
<td><strong>62</strong></td>
<td></td>
</tr>
</tbody>
</table>

Based on the confusion matrix, it is necessary to know that the green table indicates the suitability of the field validation sample data with the processing data. In contrast, the orange color indicates an error (discrepancy). From the confusion matrix, the results of the accuracy test calculations include the user’s accuracy the producer’s accuracy and the overall accuracy list in Table 4.
Table 4. User, producer and overall accuracy.

<table>
<thead>
<tr>
<th>Varieties</th>
<th>User’s accuracy</th>
<th>Producer’s accuracy</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpari 32 HDB</td>
<td>85.71%</td>
<td>88.24%</td>
<td>85.48%</td>
</tr>
<tr>
<td>Other varieties</td>
<td>85.19%</td>
<td>82.13%</td>
<td></td>
</tr>
</tbody>
</table>

From the results of accuracy test calculations based on the confusion matrix, user accuracy for the Inpari 32 HDB variety was 85.71% and for other varieties was 85.71%. Then the producer’s accuracy was 85.71% for the Inpari 32 HDB variety and 85.19% for the other varieties. Then the overall accuracy (total accuracy) was obtained at 85.48%. Through the United States Geological Survey (USGS), it has been determined that the lowest level of classification or interpretation accuracy using remote sensing is less than 85% [26], which means the overall accuracy is acceptable. Then a Kappa analysis was carried out, and the results are shown in Table 5. The kappa accuracy was 70.6% in accordance with the level of agreement carried out by Landis and Koch [27], that 70.6% is included in the moderate to high level of confidence, which is close to 80%.

Table 5. Kappa accuracy result.

<table>
<thead>
<tr>
<th>Kappa accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.6%</td>
</tr>
</tbody>
</table>

4. Conclusion

It is known that the rice varieties planted in Karangjati Sub District are the Inpari 32 HDB, Ciherang, and Cibogo varieties. In addition, based on farmer sources, there are also MR varieties planted, but not many. The dominant variety in Karangjati Sub District, Ngawi Regency, is Inpari 32 HDB. Apart from that, from the analysis of the NDVI and NDWI indices, it is known that the generative phase of rice fields in Karangjati District is around June and October 2022 and mid-January to mid-February 2023. Based on the map, the distribution of each endmember rice variety in Karangjati Sub District is spread evenly in each region. There are some rice fields that are homogeneous, and there are also those that are heterogeneous.

From the LSU results, it is known that the distribution of rice variety endmembers in Karangjati Sub District for the Inpari 32 HDB variety has a total area of 1095 ha with a range of dominant endmember values of 0.510–0.941, or 51%–94.1%, while other varieties (Cibogo, Ciherang, etc.) have a total area of 1184 ha with a non-dominant endmember value range of 0–0.25, or 0%–25%. Based on the range of endmember values of each rice variety in each Sentinel-2 image and the area of each variety, it is known that the variety that dominates at each pixel in Karangjati Sub District is the Inpari 32 HDB Variety.

The accuracy of the results of this research is acceptable, with the following details: user accuracy obtained for the Inpari 32 HDB variety was 85.71%, and for other varieties it was 85.71%. Then the accuracy of the producers for the Inpari 32 HDB variety was 85.71%, and for other varieties it was 85.19%. The overall accuracy and kappa accuracy were obtained at 85.48% and 70.6%, respectively.
Author contributions: Conceptualization, LMJ; methodology, LMJ and HS; software, MRK; validation, MRK formal analysis, MRK; investigation, MRK, LMJ and HS; resources, LMJ; data curation, LMJ; writing—original draft preparation, MRK; writing—review and editing, LMJ; visualization, MRK; supervision, LMJ and HS; project administration, LMJ; funding acquisition, LMJ. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to thank PT. Aria Agri Indonesia who has assisted in providing supporting data in this research.

Conflict of interest: The authors declare no conflict of interest.

References