Classification of rice growth stage based on convolutional neural network

R Kusumaningrum1, W Satriaji1, S N Endah1, Y Prasetyo2, A Sukmono2

1Department of Informatics, Universitas Diponegoro, Jl. Prof. Soedarto, SH, Tembalang, Semarang, 50275
2Department of Geodetic Engineering, Universitas Diponegoro, Jl. Prof. Soedarto, SH, Tembalang, Semarang, 50275

Corresponding author: retno@live.undip.ac.id

Abstract. The proposed method of rice growth classification model based on Convolutional Neural Network (CNN) which had implemented towards LANDSAT images gives the highest accuracy value of 83.4% with the following parameters including batch size 32, drop out 0.5 and band 432. The batch size value is inversely proportional to the level of accuracy obtained, which means the greater the batch size value, the smaller the average level of accuracy obtained, whereas there is no correlation between the change in the drop out value and the accuracy value and in general the best accuracy value in the drop out value is 0.5.

1. Introduction

Rice (Oryza sativa) is a kind of foods derived from the grass family. In Indonesia, rice is the leading food of most Indonesians. This makes the need for rice in Indonesia continues to increase as the population grows. However, this condition is not matched by increasing rice production. Hence, maintaining the stability of rice production in Indonesia becomes an essential issue for further study. One of the solutions that can be implemented is by applying precision agriculture (PA) in rice sector production. PA is a technology that enhances farming techniques, starting from the pre-planting, during in-season growth, through harvest and post-harvest stages [1]–[3]. One of the critical goals in PA is improving the crop yield production [4]. Therefore, monitoring the crop during in-season growth needs to be developed in the PA.

In addition, PA technologies include Remote Sensing (RS), Global Positioning System (GPS), Geographical Information System (GIS), Soil Testing, Yield Monitors, and Variable Rate Technology [5]. Remote Sensing is a science or art of getting information about an object, area or phenomenon through the analysis of data obtained without the need to make direct contact with the objects, area, or phenomenon themselves [6]. The utilization of RS for PA is taking information about the attributes of the physical, chemical and biological characteristics of the earth's surface through the image of the RS. Moreover, there are other advantages of RS image in PA, namely, remote sensing images are capable of presenting the earth's surface in the form of up-to-date and reliable objects. Therefore, this study aims to propose a novel PA technology in rice sector for during-in season growth, mainly, a rice growth phase-detection for an area using remote sensing images as the reference.

The rest of this paper consists of 4 sections, including literature review (section 2), research method (section 3), result and analysis (section 4), and conclusion (section 5). Related works discuss the latest research on the detection of the rice growth phase and LANDSAT imagery. Section 3 describes the research method in detail and its result analysis is explained in section 4. The whole study is concluded in section 5.
2. Literature review

2.1. Rice growth phase detection

There are 3 (three) growth phases in rice cultivation, namely vegetative phases, reproductive phases and ripening phases [8]. Each phase has a different duration. The vegetative phase is the most specialized phase where the duration range for this phase depends on the type of rice variety. For example, IR64 rice varieties that mature in 110 days have a 45-day vegetative phase, while varieties in IR8 that mature in 130 days have a 65-day quantitative phase. Furthermore, the reproductive phase has a duration range of approximately 35 days and the ripening phase has a duration of approximately 30 days. In addition to these three phases, there is also a resting phase in rice cultivation or commonly known as a fallow stage. The three growth phases consist of 9 (nine) different stages. These stages have sequence details as described in Table 1.

| Phase | Stage in respective phase |
|-------|---------------------------|
| Fallow Stage | 1 |
| | 2 |
| | 3 |
| Vegetative | Seedling |
| | Tillering |
| | Stem Elongation |
| Reproductive | Panicle Initiation |
| | Booting Stage |
| | Flowering Stage |
| Ripening | Milk Stage |
| | Dough Stage |
| | Mature Stage |

Table 1. Rice growth phase

There are several studies about the detection of the rice growth stage based on various remote sensing images, such as SAR image [7], MODIS image [8]–[10], PiSAR-L2 [11], and Hyperspectral image [12]. Other studies about the detection of the rice growth stage also have been conducted based on a non-remote sensing image, i.e. research-based on drone image [13], or combination between remote sensing...
image (LANDSAT 8) and Google Earth [14]. Those studies implemented various methods, such as Naïve Bayes and Linear SVM [12], SVM using Kernel RBF and Kernel-Based Regularized [8], Fuzzy Model [11], and Extreme Learning Machine [8], [10], [15]. Most of those studies implement classical machine learning, in which the classification performance depends on the precise selected features. In other words, classification performance depends on the success of the feature engineering process. However, feature engineering has weaknesses, namely expensive and requires expertise [16]. On the other hand, Deep Convolutional Neural Networks (CNN) have effectively replaced the feature engineering approaches [17], since the deep model can learn implicit features interaction [18]. Therefore, this study proposes the use of CNN to automatically classify the rice growth phases based on remote sensing images.

2.2. Remote sensing (RS)
Remote Sensing (RS) is a technology commonly used in PA. The extracted information in RS is obtained from the measurement of the number of reflections, emissions, and the spread of electromagnetic radiation from the Earth's surface, and the types of functions of wavelength, angle, wave polarization, phase, location, and time.

The used RS images in this study are images from the Landsat-8 satellite. Landsat-8, or what was formerly called Landsat Data Continuity Mission (LDCM), is an advanced technology from Landsat-7 produced by the collaboration between National Aeronautics and Space Administration (NASA) and U.S. Geological Survey (USGS). According to NASA, the goal of Landsat-8 is to expand the ability to detect and quantitatively characterize changes on the Earth's surface so that changes caused by nature and those caused by humans can be detected and differentiated [19]. Landsat-8 consists of 11 multispectral image bands as shown in Table 2.

| Bands                      | Wavelength (micrometers) | Spatial Resolution (meters) |
|----------------------------|--------------------------|-----------------------------|
| Band 1 – coastal aerosol   | 0.43-0.45                | 30                          |
| Band 2 – blue              | 0.45-0.51                | 30                          |
| Band 3 – green             | 0.53-0.59                | 30                          |
| Band 4 – red               | 0.64-0.67                | 30                          |
| Band 5 – near-infrared (NIR)| 0.85-0.88              | 30                          |
| Band 6 – SWIR 1            | 1.57-1.65                | 30                          |
| Band 7 – SWIR 2            | 2.11-2.29                | 30                          |
| Band 8 – panchromatic      | 0.50-0.68                | 15                          |
| Band 9 – cirrus            | 1.36-1.38                | 30                          |
| Band 10 – Thermal Infrared 1| 10.60-11.19            | 100                         |
| Band 11 – Thermal Infrared 2| 11.50-12.51            | 100                         |

2.3. Convolution neural network (CNN)
Convolutional Neural Network (CNN) is one of the deep learning methods that are good for image recognition because of CNN's ability to imitate image recognition in the visual cortex of humans. CNN is a development of Multilayer Perceptron (MLP) which is designed to process two-dimensional data. The way CNN works has something in common with MLP, but every neuron in CNN is represented in two-dimensional form, unlike MLP where each neuron is only one-dimensional in size.

This gives effect in the form of data propagated on the CNN network which is two-dimensional data, so that linear operations and weight parameters on CNN are different from MLP. Linear operation on CNN uses convolution and weight operations on four-dimensional CNN.
According to [20], the type of layer on CNN is divided into two, including:

a. The feature extraction layer, consisting of two layers located at the beginning of the architecture which is composed of several layers and each layer is composed of neurons connected to the local area (local region) of the previous layer. The two layers in this feature's extraction layer are the first type layer called the convolution layer and the second layer is the pooling layer. Each layer has an activation function. Its position is alternating between the first type and the second type. This layer accepts image input directly and processes it until it produces an output in the form of a vector to be processed in the next layer.

b. The classification layer composed of several layers and each layer is composed of fully connected neurons with other layers. This layer accepts input from the extraction feature layer output in the form of a vector then transformed like Multi Neural Networks with the addition of several hidden layers. The output is class scoring for classification.

The following is an explanation of the feature extraction layer and the classification layer:

a. Convolution layer
   The output size of the convolutional layer is influenced by the form of the input as well as the kernel size, padding and strides [21]. The convolutional layer consists of input $I$, kernel filter $K$, and bias $b$ and output in the form of a feature map. A feature map will be generated when this filter is shifted to all parts of the image. Each shift will do dot operations between the input and the value of the filter. Figure 1 shows the convolution process on CNN.

\[
(l * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} K_{m,n} I_{i+m,j+n} + b
\]

(2)

The activation function used in the convolution layer is ReLU. The ReLu activation function is an activation function that has a range between 0 to infinity.

b. Pooling layer
   The pooling layer reduces the size of feature maps to summarize subregions such as using the max-pooling function or average pooling [21]. Pooling works by sliding the window on all feature maps. The following is an example of max-pooling operation as seen in Figure 2.
c. Fully connected layer

After the process of convolutional layer and pooling layer is done, the fully connected layer is continued. Before being processed as input to the fully connected layer, the results of feature maps in the form of multi-dimensional arrays are converted into vector shapes. A fully connected layer like MLP has a hidden layer, activation function, output layer, and loss function. The activation function used in the output is softmax. The following is the softmax activation function:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{k} e^{x_j}}$$

(3)

The softmax function will calculate the probability of each target class for all possible target classes. The output value ranges from 0 to 1 and the sum of all probabilities will be equal to 1. If used for multi-classification models, this function will return the probability of each class and the target class will have a high probability.

The loss function is used to measure the performance of the neural network in predicting targets. The way the loss function works is to compare the prediction results from the output layer with the target. The loss function used in this study is categorical cross-entropy.

Deep learning algorithms have many numbers of parameters and tend to be slow. The size of the network formed by deep learning algorithms causes overfitting problems. One way to overcome the problem of overfitting is using drop out techniques [22]. The main idea of using drop out is to randomly choose units not to participate in the training process in subsequent neurons, in other words, discarded [22].

3. Research method

3.1. CNN network architecture

In the previous chapter, the author discussed the layers that compose CNN. This chapter will discuss the network architecture of combining CNN layers. Details of the developed network architecture can be seen in Figure 3.
Based on Figure 3, the CNN network architecture developed in this study consists of one input, 6 convolutional layers, 3 max-pool layers, 3 fully connected layers. The first convolution layer accepts input with a dimension of 30×30×3. The second convolution layer accepts input coming from the first convolution layer with a dimension of 30×30×32. In the convolution layer using a kernel size with a 3×3 dimension and its activation function ReLU, it doesn't use stride and padding. Next, the first layer pooling with a 2×2 size pool that produces output with a dimension of 15×15×32. Then the dropout technique is performed with a value of a combination of 0.25; 0.5; 0.75.

Convolution of the third- and fourth-layers results in output dimension of 15×15×64 with kernel size with a 3×3 dimension and the activation function is ReLU, then the second layer pooling with pool size as in the first layer which produces an output of 7×7×64. The results of the second pooling layer are passed in the dropout.

The fifth- and sixth-layer convolution accepts input with 7×7×128 dimension with kernel size and activates the function as in the previous convolution layer. Then it is passed to the pooling layer with the 2×2 pool size so that it produces a matrix with a size of 3×3×128 and is followed by a dropout.

Before being processed in a fully connected layer, a matrix with a dimension of 3×3×128 is converted to vector shape with the number of neurons as much as 1152. This architecture uses 2 fully connected layers with ReLU activation and 1 fully connected layer with softmax activation which produces 10 output nodes for classification.

Adam Optimizer is used to find convergent weights between layers so that the obtained weights produce the desired output, with the learning rate of 0.0001 and decay of 1×10^{-6}. The epoch used by the author was 10,000, with 3 batch combinations 32, 64, 128.
3.2. Training and evaluation
The data is trained in the form of a matrix of 30×30×3 with a total data of 843. The data is divided into
2, namely training and testing data with a ratio of 1:3 so that the total training data is 618 and testing
data is 265. Labeling in this data uses 10 classes as explained as rice growth stages in Table 1. The label
is changed to categorical form with the label 0-9. The data is trained using network architecture as
described in Section 3.1. The evaluation method used by the author is metric accuracy and loss of
categorical cross-entropy function.

4. Implementation and results

4.1. Experiment and dataset environment
The environment for developing a classification model based on CNN (Convolutional Neural Network)
in the rice growth phase uses the Google Collaboratory and the Python programming language. The
dataset used in this study is a remote sensing image (LANDSAT). The dataset was obtained from field
surveys in various rice fields in Kendal, Pekalongan, and Demak Regencies. The dataset obtained is
then divided into 2, namely:

a. Development Dataset
The datasets criteria taken for the category of the taking of rice fields coordinates that fall into the
dataset development category are the area of rice fields with a homogeneous rice growth phase
covering 900 m × 900 m. The number of development datasets used is 843. Table 3 is the distribution
of the number of paddy fields for each rice growth phase.

| Rice Growth Phase | Total of Sample Data |
|-------------------|----------------------|
| 0                 | 73                   |
| 1                 | 73                   |
| 2                 | 77                   |
| 3                 | 95                   |
| 4                 | 72                   |
| 5                 | 48                   |
| 6                 | 100                  |
| 7                 | 104                  |
| 8                 | 97                   |
| 9                 | 104                  |
| **Total:** 843    |                      |

b. Validation Dataset
The criteria that will be taken for the taking of the rice field area coordinates included in the
development dataset category is the rice field area with heterogeneous rice growth phase with area
varying above 900 m × 900 m.

4.2. Experiment and analysis results
The parameters tested in this study are bands from an image file, batch size, drop out. Bands from image
files are Band 543 and Band 432; batch size is 32, 64, and 128; and drop out is 0.25, 0.5, and 0.75. All
parameter combinations are tested to get the best accuracy. The best accuracy will be used to develop
applications for rice crop-growing profilers. Input data used in this study is remote sensing (LANDSAT)
imagery. This study uses 2 metrics (accuracy and lost function) and confusion matrix. The following is
a summary of the results of the performance evaluation as seen in Table 4.
Table 4. Performance evaluation of the rice growth phase classification model using CNN.

| Batch Size | Drop Out | Accuracy          |
|------------|---------|-------------------|
|            |         | Band 432 | Band 543 |
| 128        | 0.25    | 0.423    | 0.563    |
| 128        | 0.5     | 0.8      | 0.783    |
| 128        | 0.75    | 0.219    | 0.179    |
| 64         | 0.25    | 0.453    | 0.574    |
| 64         | 0.5     | 0.8      | 0.806    |
| 64         | 0.75    | 0.223    | 0.183    |
| 32         | 0.25    | 0.491    | 0.65     |
| 32         | 0.5     | 0.834    | 0.806    |
| 32         | 0.75    | 0.215    | 0.194    |

Based on Table 4, it can be seen that the best accuracy value is obtained at a batch size 32, drop out 0.5 and band 432, with an accuracy value of 83.4%. While the highest accuracy for the 543 bands is obtained in the same batch size and drop out size pairs, and 64 size batch size and 0.5 dropout size with an accuracy value of 80.6%.

Based on Figure 4, it can be seen that the batch size value is inversely proportional to the level of accuracy obtained, which means the greater the batch size value, the smaller the average level of accuracy obtained. This is because the use of large batches tends to blend with the sharp minimizer of the training function. This minimizer is characterized by a large positive eigenvalue in $\nabla^2 f(x)$ and tends to generalize poorly. In contrast, small-batch methods merge with a flat minimizer which is characterized by a small positive eigenvalue of $\nabla^2 f(x)$.

Based on Figure 5, it can be seen that there is no correlation between changes in the value of drop out and the value of accuracy. The drop out process will turn off some neurons in the neural network randomly so that we have a rare network that greatly reduces the possibility of overfitting. The drop out value is generally made in the range of 0.3 to 1.0, which means that when the drop out value is 1.0, there is no single neuron that is turned off. This is because the selection of neurons that are turned off is done randomly so that the value of the batch size does not have a pattern that shows a correlation between the change in the dropout value and the accuracy value.

![Figure 4. Effect of batch size changes on accuracy value.](image-url)
5. Conclusion
Based on the results of the discussion in the previous chapter, the following conclusions can be drawn as follows:

1. The classification model of rice plant growth phase using Convolutional Neural Network obtained with parameters including batch size 32, drop out 0.5 and band 432, with an accuracy value of 83.4%.
2. The batch size value is inversely proportional to the level of accuracy obtained, which means the greater the batch size value, the smaller the average level of accuracy obtained.
3. There is no correlation between the change in the dropout value and the accuracy value and in general, the best accuracy value in the drop out value is 0.5.

Acknowledgments
The authors would like to acknowledge the research funding supported by Universitas Diponegoro under the grant of applied research for college flagship No.101-142/UNJ.P4.3/PP/2019 – the Year 2019.

References
[1] Abdullahi H S, Sheriff R E, and Mahieddine F 2017 Proceeding of The Seventh International Conference on Innovative Computing Technology (INTECH)
[2] Bongiovanni R and Lafayette W 2004 Precision Agriculture and Sustainability pp. 359–387
[3] Stafford J V 2000 J. Agric. Eng. Res. vol. 76 pp. 267–275
[4] Chlingaryan A, S Sukkarieh and B Whelan 2018 Comput. Electron. Agric. vol 151 no. June pp. 61–69
[5] Venkatalakshmi B and Devi P G 2017 IJRET Int. J. Res. Eng. Technol. vol. 3 no. 7 pp. 849–852
[6] Lillesand T, Kiefer R W, and Chipman J 2015 Remote Sensing and Image Interpretation 7th Editio. Wiley
[7] Yuzugullu O, Marelli S, Erten E, Sudret B, and Hajnsek I 2017 Remote Sens. vol. 9 no. 5 pp. 1–20
[8] Mulyono S, Harisno H, Amri M, Fanany M I, and T Basaruddin 2015 Kernel-Based Regularized Learning from Time-Invariant Detection of Paddy Growth Stages from MODIS Data, in Lecture Notes in Computer Science no. June N T Nguyen, B Trawiński, and R Kosala, Eds. Springer Cham pp. 273–283.
[9] Mulyono S, Sadly M, Fanany M I, and Basaruddin T 2015 J. Theor. Appl. Inf. Technol. vol. 71 no. 3 pp. 430–439
[10] Mulyono S, Pianto T A, Fanany M I, and T Basaruddin 2013 International Conference on Advanced Computer Science and Information Systems, ICACSIS no. September pp. 309–314.
[11] Mulyono S, Sumargana, I Fauziyah, and E Kustiyanto 2013 34th Asian Conf. Remote Sens. ACRS vol. 3 no. December pp. 2530–2537
[12] Halim H, Isa S M, and Mulyono S 2016 IEEE Region 10 Symposium (TENSYMP) vol. 2 pp. 167–172
[13] Marsujitullah M, Zainuddin Z, Manjang S, and Wijaya A S 2019 IOP Conf. Ser. J. Phys. vol. 1198 no. 092001 pp. 1–7
[14] Dong J et al. 2016 Remote Sens. Environ. vol. 185 pp. 142–154
[15] Suhandono N, Masiyanti F, and Fanany M I 2013 An Extreme Learning Machine Model for Growth Stages Classification of Rice Plants from Hyperspectral Images Subdistrict Indramayu no. October
[16] Liu B, Yu J, Tang R, Guo H, Y Chen, and Y Zhang 2019 The Web Conference 2019-Proceedings of the World Wide Web Conference pp. 1119–1129
[17] Pintea S L, Mettes P S, Van Gemert J C, and Smeulders A W M 2016 Proc. - Int. Conf. Image Process. ICIP pp. 196–200
[18] Cheng H-T et al. 2016 Proceedings of the 1st Workshop on Deep Learning for Recommender Systems pp. 7–10
[19] Loveland T R and Irons J R 2016 Remote Sens. Environ. vol. 185 pp. 1–6
[20] Zafar M and Setiyono B 2016 J. Sains Dan Seni ITS vol. 5 no. 2 pp. 72–77
[21] Dumoulin V and Visin F 2016 A guide to convolution arithmetic for deep learning pp. 1–31
[22] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and R Salakhutdinov 2014 J. Mach. Learn. Res.