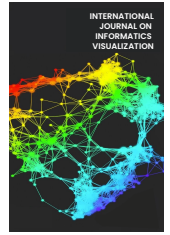




INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Determining the Rice Seeds Quality Using Convolutional Neural Network

Sidiq Syamsul Hidayat ^{a,*}, Dwi Rahmawati ^b, Muhamad Cahyo Ardi Prabowo ^a, Liliek Triyono ^a,
Farika Tono Putri ^a

^a Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang, 50275, Indonesia

^b Department of Seed Production Techniques, Politeknik Negeri Jember, Jember, 68121, Indonesia

Corresponding author: *sidiqsh@polines.ac.id

Abstract— Seed inspection is crucial for plant nurseries and farmers as it ensures seed quality when growing seedlings. It is traditionally accomplished by expert inspectors filtering samples manually, but there are some challenges, such as cost, accuracy, and large numbers. Speed and accuracy were the main conditions for increasing agricultural productivity. Machine learning is a sub-science of Artificial Intelligence that can be applied in research on the classification of rice seed quality. The pipeline of a machine learning system is dataset collection, training, validation, and testing. Model making begins with taking data on the characteristics of rice seeds based on physical parameters in the form of seed shape and color. The dataset used is two thousand images divided into two categories, namely superior seeds and non-superior seeds. Training and Validation was conducted using the Convolutional Neural Network (CNN) algorithm with the concept of cross-validation on Google Collaboratory notebooks. The ratio split of train data and validation data in modeling from a dataset is 80:20. The result of the model formed is a model with the development of a Deep Convolutional Neural Network (Deep CNN) that can classify the digital image data of rice seeds from the results of data calls uploaded into the system. The results of the experiment conducted on 30 test data can be analyzed so that the system can classify superior and non-superior seeds with a precision value of 93% and a recall of 95%.

Keywords— Artificial intelligence; machine learning; CNN; classification of rice seeds; deep CNN.

Manuscript received 31 Aug. 2022; revised 5 Oct. 2022; accepted 8 Nov. 2022. Date of publication 30 Jun. 2023.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Rice is one of the leading food products in agriculture. In the Asian region, especially in Indonesia, rice is one of the basic ingredients of basic food that are the main and important. Most people meet their food needs by consuming rice made from raw materials [1]. Inspecting rice (*Oryza sativa*) seed variety is a critical procedure for quality assessment in the arable sector [2]. The challenge for professionals and farmers at seedling propagation stations is ensuring that all batch seeds are from the same variety. By introducing weeds and off-types into the crop, varietal contamination can affect rice seed yields.

Several problems can be caused by contamination, including variety impurities, rice mutations, and crossbreeding, which can lead to poor-quality crops. Rice experts have conducted traditional techniques for examining contamination in breeding seeds. The differences between types of paddy seeds are sometimes hard to distinguish

because they are so small. To decide, the experts consider many parts of the seed, such as its shape, texture, and color. The researchers examined rice seeds from localities and classified them according to their type. First, they examine seeds using tools such as a magnifying glass, a flashlight, and forceps to determine whether they are the same type. Later, they seek out seeds that exhibit different physical characteristics and are contaminated seeds. With the limitations of being human, a large number of seed inspectors take quite a long time in the process because it is difficult for the human eyes to find small differences in one seed among many seed samples [3].

Technology in the industrial era 4.0 is present and provides solutions in various fields, not least in agriculture. The development of technology to be applied in the management of rice seeds quality standard testing is needed to make an impact in the form of efficiency of rice seeds quality standard assessment on a large scale [4]. Another impact is to encourage the improvement of rice processing to produce

products with export quality. Technology in the industrial era 4.0 is present and provides solutions in various fields, not only agriculture. The development of technology to be applied in the management of rice seeds quality standard testing is needed to make an impact in the form of efficiency of rice quality standard assessment on a large scale. Another impact is to encourage the improvement of rice processing to produce products with export quality.

Computer vision has been widely applied in various fields in the past decade. Deep learning has replaced statistical methods for tasks such as object detection and image recognition in computer vision because it offers greater accuracy. Computer scientists can use technology in various fields to quickly develop tasks. Traditional machine learning methods need to engineer features individually while automatically learning features from the given data. This model can handle very similar variability and deviations in data. Although deep learning technology is rather complicated, it has much potential. The network structure is large, and the training data and high-performance computing resources are time-consuming, making it complex. To compare classification performance between traditional machine vision and deep learning methods, we tested various classification methods of rice cultivars.

In this study, rice quality was identified by one method. The method is supervised learning. The supervised learning method was designed to create a computer that could have intelligence in determining rice seeds classification by studying patterns in data. This system is designed to classify rice seeds that amount to one grain into two types: superior and non-superior.

The analysis was conducted on the results of making rice seed quality detection systems using supervised learning methods. With the development and analysis of the system made, this technology is expected to help humans overcome the assessment of rice seed quality has a high level of complexity and requires a high level of accuracy to produce an accurate assessment in the field of quality of agricultural products.

A Machine vision in agriculture applications related to rice quality inspection and grading has been reviewed and summarized [3]–[5]. There are various works related to the field of rice quality inspection. Rice quality measurement can be performed on milled rice and paddy seeds, depending on the purpose. Several studies were presented in the following work that examined the quality inspection methods for milled rice grains and polished rice. Analyzed five rice varieties with quite different shapes and colors, the classification of various rice grain varieties is made using image processing techniques and algorithms [6]. In some previous works, seeds with defects mixed with grains were analyzed [7], [8].

Various defects were studied, including broken, chalky, damaged seeds and improper elements. Detecting and classifying defective grains allowed for estimating the purity of rice grains. Image processing technique with a k-NN classifier evaluated three classes (30 images for each class) [2]. Some other research emphasized detecting the chalkiness that appeared in the grain [9]. Chalky grains have a partially opaque or milky white kernel. The degree of chalkiness is one of the most important evaluation indicators. Chalky rice grains break during milling, which will affect their taste.

Many studies have proposed a rice seed classification technique analyzing information from a hyperspectral imaging system [1], [10]–[13]. In hyperspectral imaging, a wide range of electromagnetic spectrums can be seen with a higher degree of spatial detail. Suitable for analyzing surfaces of materials, it can be used for surface analysis. Additionally, many of them developed deep learning techniques (CNNs) in addition to traditional classification methods. Hyperspectral image data from a near-infrared camera was used to classify six common rice seed varieties and evaluated 108 seeds for each variety and 648 seeds across all varieties. The classification used SVM and a Random Forest classification technique [14]. Rice seeds' accuracy increased to 84% when spectral and shape-based properties were combined, compared to 74% when only visual properties were used.

The Rice Seed Germination Evaluation System (RSGES) was developed to evaluate rice seed images to predict germination using digital image processing techniques and artificial neural networks. Digital images are taken with a digital camera. RSGES consists of six main processing modules: 1) image acquisition, 2) image pre-processing, 3) feature extraction, 4) germination evaluation, 5) result presentation, and 6) germination verification [15]. Accordingly, many researchers have proposed grain classification (sometimes with localization) methods based on convolutional neural networks (CNNs) for grain purity inspection by an image. However, those papers are necessary to have a large number of labeling that was too expensive to be manually collected [11]. An article describes how to determine rice seed varieties using image processing techniques and artificial neural networks (ANNs) based on extracted color features. The experiment used 200 individual seed images of two Malaysian rice seed varieties, MR 219 and MR 269 [16]. The recognition process is crucial for ensuring the purity of rice varieties. To characterize rice seed images, we propose to use the histogram of oriented gradient (HOG) descriptor. Since the size of the image is totally random, the features extracted by HOG cannot be used directly by the classifier due to the different dimensions [17].

An algorithm based on deep learning is developed in the proposed solution. Annotations are made using the watershed algorithm. Alignment is auto-aligned based on major axis orientation, and image enhancement is applied using contrast-limited adaptive histogram equalization (CLAHE). Then, the mask region-based convolutional neural network (R-CNN) is trained to localize and classify rice grains in an input image [18]. For various kinds of lung abnormalities, including nodules and diffuse lung diseases, R-CNN is used to implement image-based CADe [19]. Segmentation of cervical cells using a mask regional convolutional neural network (Mask R-CNN) and classification using a smaller Visual Geometry Group-like Network (VGG-like Net). In mask R-CNNs, ResNet10 is used to maximize the use of spatial information and prior knowledge. We evaluate our proposed method based on the Herlev Pap Smear dataset [20], [21]. Inception modules lie between convolutions and depth wise separable convolutions on the discrete spectrum. The future of convolutional neural networks will be built on depth wise separable convolutions [22], [23].

Augmenting training datasets in order to build better Deep Learning models is known as Data Augmentation [24]. An

effective CNN-based dual-phase method has been proposed for small, heterogeneous datasets of rice grain diseases. To determine the variety of rice grains, multilayer perceptron and neuro-fuzzy neural networks were used [25].

The study proposes a hyperspectral method using LLRM for rice seed purity identification, which improves recognition accuracy and selects feature wavelength bands stably, with a range of 91.67% to 100% and an average of 98.47%. The study demonstrates the feasibility of developing hyperspectral technology for seed purity identification, but cost and optimization remain limitations[26]. This study proposes an automated machine vision solution using morphological and spectral features for rice seed classification. The proposed approach achieved an average F1 score of 85.65% on a dataset of 8640 seed samples corresponding to 90 different rice seed varieties, demonstrating the potential of combining RGB and hyperspectral imaging for accurate classification [27].

The study evaluated MV techniques for classifying six Asian rice varieties using digital field images captured by a cell phone camera. Binary, Histogram, and Texture features were extracted from non-overlapping regions of interest. After optimization, LMT Tree achieved the highest accuracy of 97.4%, with a maximum overall accuracy (MOA) of 97.4%[28]. This study uses a deep learning technique called MobileNetV2 to classify 14 different types of seeds accurately. Deep learning has shown significant potential for agricultural applications, including crop disease identification and seed image analysis[29]. Rice is a vital source of food for over half the world's population, and its production is predicted to increase in the coming years. Rice is packed with essential nutrients such as vitamins, minerals, and antioxidants, making it superior to other staple foods. This review discusses global rice production, varieties, consumption, nutritional values, and environmental impacts. It also includes a new method of paddy storage, drying, and grading rice [30].

II. MATERIALS AND METHOD

The research on rice quality identification was conducted on a laptop running Windows 10 and powered by an Intel Core i7 CPU. A Xiaomi Redmi Note 8 smartphone camera with a 2MP resolution macro lens is used to retrieve rice image data. The software used to design the system is Google Collaboratory, TensorFlow library, and Python programming language. The dataset retrieval is done in a Studio Box, which has been adjusted for lighting and has a white background in order to distinguish it from the object being studied. The type of rice seeds used is Situbagendit seeds, with the category of superior seeds and non-superior seeds.

The research methodology applied in rice seed quality identification uses supervised learning with convolution neural network architecture. Supervised learning methods with convolution neural network techniques are used to predict the quality of rice with a limit of 1 grain of rice seed. The input for this processing is the rice seed image, and the output from this processing is the identified rice seed category information.

A. Rice Seed Classification System

The working diagram of the rice seed quality identification system based on Figure 1 is divided into two stages. The 1st

stage, namely the Model Training Phase, is to create a database and system model based on the classification of rice seed parameters by taking samples of rice seed images. The 2nd stage, namely the Test Data Test Phase, conducts a test for the detection of rice seed quality using image data that is not known seed quality by integrating the test data image with the model that has been made to be detected and known the classification results.

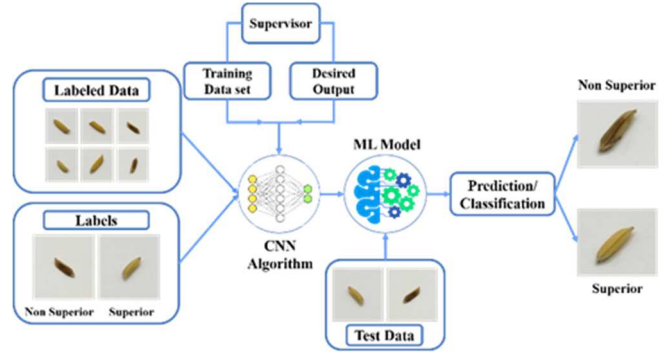


Fig. 1 System classification of rice seeds – Overview

B. Image Acquisition

An image acquisition system retrieves rice seed images for training and testing rice identification systems. Figure 2 shows the scheme for taking rice seed images. A three-foot camera brace is used for rice image-taking aids. Objects and the camera are always at the same distance for each rice seed, with a white background image.



Fig. 2 Techniques for collecting image data

C. Dataset

A dataset is a set of data that is an important part of creating a machine-learning model. The stages in creating a dataset are data collecting and data cleaning. In the case of data collecting, finding the right dataset is one of the important steps in a machine learning project. At this stage, the local seed variety used is the Situbagendit seed type. The rice seed dataset is divided into two classifications: superior and non-superior. Each column represents one class of rice seed varieties. Data cleaning is carried out to improve data quality which will impact the work productivity of the system. Inaccurate data can adversely affect the accuracy and performance of a model. In the data cleaning process, unnecessary data and information are discarded to obtain quality data. Accurate and quality data will have a positive effect on the decision-making process. The number of datasets used in this study amounted to 2000 datasets, divided

into 1000 superior seed datasets and 1000 non-superior seed datasets.



Fig. 3 A sample of the rice seeds dataset

D. Supervised Learning

In machine learning, supervised learning uses data training to create classification or regression systems.

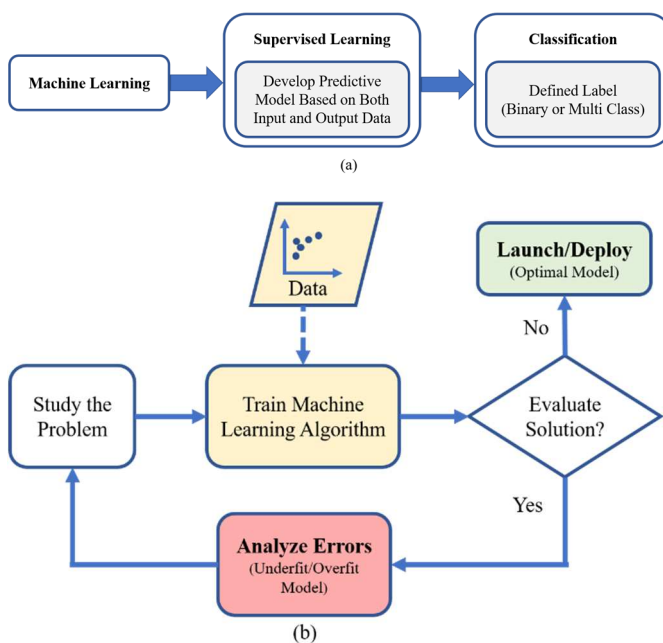


Fig. 4 (a) Supervised learning, (b) Machine learning algorithms

Figure 4 (a) Supervised Learning. The machine learning process develops predictive models based on inputs and outputs. This study applies supervised learning by going into the discussion of classification for rice seed quality. Figure 4 (b) Machine Learning Algorithms shows that machine learning algorithms develop models based on data sets. Machine learning looks for specific patterns from each rice seed data set that have been grouped by rice seed size or shape category. This data set is used as training data in model development. Models formed from training data sets are further applied to predict categories of rice seeds whose identity is not yet known. The results of the system prediction will produce output in the form of rice seed quality information.

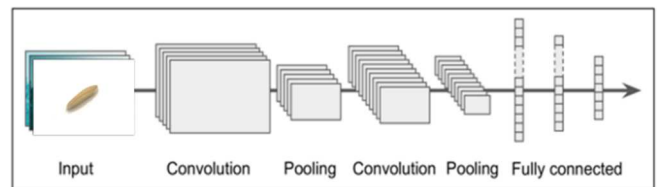


Fig. 5 Architecture of convolutional neural networks

As shown in Figure 5, the architecture of convolutional neural networks, Convolutional neural networks go through several stages [13]. Convolution neural networks are one of the supervised learning algorithms used in this study to identify rice quality based on the creation of models obtained from convolution neural network algorithms training. Processing in a convolution network includes several stages, namely:

1) *Input*: This stage is retrieving rice image data in format. Jpg is used for input which will then be detected at the convolution stage.

2) *Convolution*: This stage serves to recognize the unique attributes of an object. Networks are designed to focus on low-level features in the first hidden layer and then combine them in the next hidden layer to create higher-level features. This layer applies filters to help identify the image's features.

3) *Pooling*: This stage forms a secondary feature extraction effect, can reduce the dimensions of the feature map, and increase the durability of feature extraction on the image. The result of the pooling process is passed on in the labeling connection.

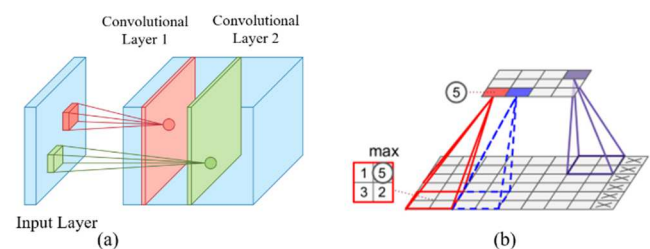


Fig. 6 (a) Process of convolution, (b) Process of pooling

E. Data Pre-processing



Fig. 7 Pre-processing flowchart

1) *Images—Channels and Sizes*: Images come in different shapes and sizes. Considering all these variations, any image data needs to be pre-processed. Digital images are encoded in RGB format. The first step in data pre-processing is to create an image with the same dimensions.

2) *Morphological Transformations*: The term morphological transformation refers to any modification involving the shape and form of the images. The typical transformations are erosion, dilation, opening, and closing.

- **Thresholding.** A simple operation that converts all pixels with intensities above a certain threshold to ones. The pixels having a value less than the threshold are converted to zero. This results in a binary image.
- **Erosion, Dilation, Opening & Closing.** Erosion shrinks bright regions and enlarges dark regions. However, dilation works in the opposite direction, enlarging bright and shrinking dark regions. The opening is erosion, followed by dilation. In addition to removing small bright spots, the opening might connect small dark cracks. Closing is dilation followed by erosion. The closing of small cracks can eliminate dark spots and connect bright spots.

3) *Normalization:* Normalization is the most crucial step in the pre-processing part. This refers to rescaling the pixel values within a limited range. One of the reasons to do this is to help with the issue of propagating gradients.

4) *Augmentation:* To increase the amount and variance of training data, augmentation is often used in image-based deep learning tasks. When creating a model, the training set should be augmented. The model learns all variations, including flipping, rotation, cropping, translation, illumination, scaling, adding noise, etc. This significantly boosts the accuracy of the model.

F. Model Designing with Google Collaboratory

Google Collaboratory is a software owned by Google for developing artificial intelligence systems using the Python programming language. The designed system starts with creating a Google account and connecting to the Google Collaboratory Notebooks service. The model is designed by calling a dataset and entering training with a convolutional neural network algorithm. Models formed and validated according to the target will be saved for further use to perform classification testing.



Fig. 8 A machine learning software package

G. Tuning CNN Parameters

Model creation is done by dividing the percentage of datasets totaling 2000 images into 80% training data (1600 images) and 20% validation data (400 images). Use an image data generator. Parameters `rescale = 1./255`, `rotation_range = 20`, `horizontal_flip = True`, `shear_range = 0.2`, `fill_mode = 'nearest'`. The model used is sequential, with convolution stages of 1 to 4. The next stage is the flattened and dense layer neurons. The output of this process becomes the initial stage before entering the training phase. Model optimization is added with Adam's type of optimization with a learning rate of 0.001.

Data training against the designed model is carried out by determining the expected accuracy percentage. In this case study, a percentage of 91% was set to make the model stop in the training process. Epoch is the number of iteration or repetition processes in the specified model training a maximum of 50 times. Batch size is the number of images

entered in each training step, referring to 32, 64, 128, and multiples standards. Step per epoch is the number of steps to complete 1 epoch, which can be determined from the number of training data, which is 1600 data divided by a batch size of 64. Validation steps are the number of validation steps to complete 1 epoch, which can be determined from the number of validation data, which is 400 data divided by a batch size of 64.

III. RESULTS AND DISCUSSION

This supervised learning model was made using a 2000 rice seed image data set consisting of one thousand superior rice seed image data and 1000 non-superior rice seed images. The results of making a model with a data set of 2000 produce 1 type of model with a comparison ratio of 80% training data: and 20% validation data.

A. Model with Ratio 80% Training Data: 20% Validation Data

Figure 9 shows the percentage relationship of accuracy with the training iteration of a directed learning model using a dataset of 2000 rice seed images. The creation of this model applies the division of the ratio of the image data set, which is 80% of the training data: and 20% of the validation data. The number of iterations performed is indicated on the X-axis, and the resulting model accuracy values are indicated on the Y-axis. The model's accuracy in this training consists of training and test data. Based on the planned iteration process of 50 times epoch, the program has stopped at the 39th epoch because it has reached the highest accuracy target in the model. Analyzing the data obtained, the accuracy value of the model changed from 0.5 at the beginning to 0.9 at the end of the 39th iteration. Indicates there is an improvement in the accuracy value of each iteration until it reaches the expected accuracy value of 0.9.

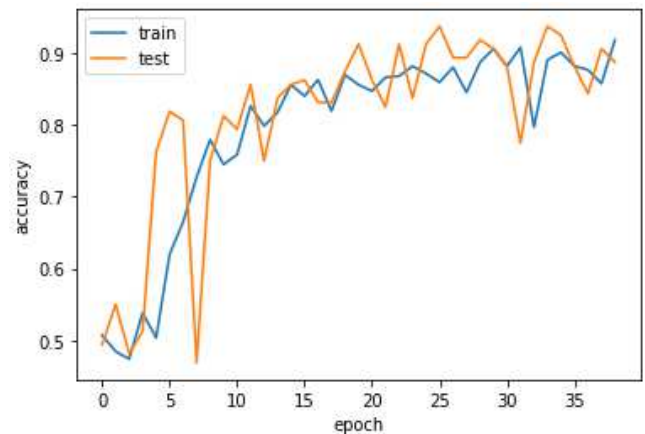


Fig. 9 A correlation between accuracy and the epoch process

Figure 10 shows the model error values from the training and validation data's training process. This figure informs the result of comparing the number of iterations shown on the X axis and the value of the error percentage of the model shown on the Y axis. Model errors in this training consist of train and validation data errors. Based on the planned iteration process of 50 times epoch, the program has already stopped at the 39th epoch because it has reached the target of the smallest error in the model. According to the data analysis, the error value

changed from the 1st iteration, which is 0.8, to the 39th iteration, which is 0.2.

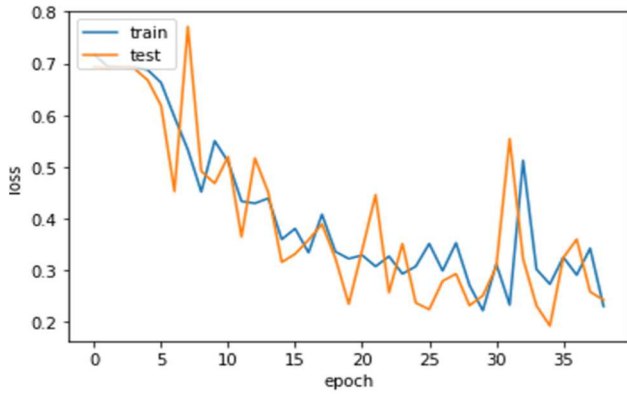


Fig. 10 A correlation between loss and the epoch process

B. Training and Validation Process of Epoch Program

A model's accuracy level can be known in value from the epoch results read at the time of model creation in Google Collaboratory. Each epoch read on the system will show the resulting model accuracy and error value. Ideally, in creating machine learning models, accuracy in each epoch will be of increasing value to either or higher. As for the model's error, the idea and the number of epochs read will go down. Figure 11 and Figure 12 show the results of the epoch process read in creating a seed identification model.

18/18 - 180s - accuracy: 0.8802 - val_accuracy: 0.8938 - 180s/epoch - 10s/step
Epoch 28/50
18/18 - 170s - accuracy: 0.8455 - val_accuracy: 0.8938 - 178s/epoch - 10s/step
Epoch 29/50
18/18 - 179s - accuracy: 0.8872 - val_accuracy: 0.9187 - 179s/epoch - 10s/step
Epoch 30/50
18/18 - 184s - accuracy: 0.9062 - val_accuracy: 0.9062 - 184s/epoch - 10s/step
Epoch 31/50
18/18 - 183s - accuracy: 0.8819 - val_accuracy: 0.8813 - 183s/epoch - 10s/step
Epoch 32/50
18/18 - 181s - accuracy: 0.9080 - val_accuracy: 0.7750 - 181s/epoch - 10s/step
Epoch 33/50
18/18 - 181s - accuracy: 0.7969 - val_accuracy: 0.8875 - 181s/epoch - 10s/step
Epoch 34/50
18/18 - 183s - accuracy: 0.8906 - val_accuracy: 0.9375 - 183s/epoch - 10s/step
Epoch 35/50
18/18 - 183s - accuracy: 0.9010 - val_accuracy: 0.9250 - 183s/epoch - 10s/step
Epoch 36/50
18/18 - 184s - accuracy: 0.8819 - val_accuracy: 0.8813 - 184s/epoch - 10s/step
Epoch 37/50
18/18 - 186s - accuracy: 0.8767 - val_accuracy: 0.8438 - 186s/epoch - 10s/step
Epoch 38/50
18/18 - 184s - accuracy: 0.8576 - val_accuracy: 0.9062 - 184s/epoch - 10s/step
Epoch 39/50
18/18 - 185s - accuracy: 0.9184 - val_accuracy: 0.8875 - 185s/epoch - 10s/step

Fig. 11 Percentage accuracy data from the epoch model

18/18 - 180s - loss: 0.2991 - val_loss: 0.2793 - 180s/epoch - 10s/step
Epoch 28/50
18/18 - 170s - loss: 0.3529 - val_loss: 0.2932 - 178s/epoch - 10s/step
Epoch 29/50
18/18 - 179s - loss: 0.2731 - val_loss: 0.2321 - 179s/epoch - 10s/step
Epoch 30/50
18/18 - 184s - loss: 0.2223 - val_loss: 0.2510 - 184s/epoch - 10s/step
Epoch 31/50
18/18 - 183s - loss: 0.3124 - val_loss: 0.3067 - 183s/epoch - 10s/step
Epoch 32/50
18/18 - 181s - loss: 0.2332 - val_loss: 0.5541 - 181s/epoch - 10s/step
Epoch 33/50
18/18 - 181s - loss: 0.5121 - val_loss: 0.3223 - 181s/epoch - 10s/step
Epoch 34/50
18/18 - 183s - loss: 0.3020 - val_loss: 0.2316 - 183s/epoch - 10s/step
Epoch 35/50
18/18 - 183s - loss: 0.2733 - val_loss: 0.1925 - 183s/epoch - 10s/step
Epoch 36/50
18/18 - 184s - loss: 0.3254 - val_loss: 0.3244 - 184s/epoch - 10s/step
Epoch 37/50
18/18 - 186s - loss: 0.2910 - val_loss: 0.3598 - 186s/epoch - 10s/step
Epoch 38/50
18/18 - 184s - loss: 0.3424 - val_loss: 0.2585 - 184s/epoch - 10s/step
Epoch 39/50
18/18 - 185s - loss: 0.2305 - val_loss: 0.2424 - 185s/epoch - 10s/step

Fig. 12 Percentage loss data from the epoch model

C. Supervised Learning System Prediction Test

The model formed from the training data learning process is then used to test the rice seed quality predictions on the rice seed test data. Based on the results of system testing on test data, the model's accuracy in predicting rice seed quality. A total of 60 image data were collected from the process of making test data for this study. The test data, which amounted to 60 images, consisted of 30 test data for superior rice seed and 30 test data for not-superior type. The model to be tested is a model with a data set of 2000 data.

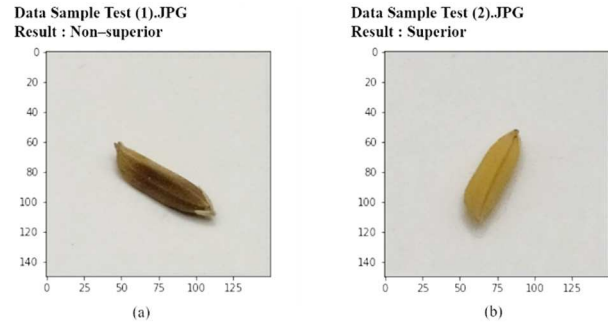


Fig. 13 (a) Non-Superior Rice seed (b) Superior Rice seed

Figure 13 shows the model prediction results on the test data, namely the image of the superior rice seed and the image of the non-superior rice seed. The system will call the image data file and process the file by predicting the image based on the image's relevance in the training data and then generating an image prediction label. The prediction system in identifying rice seed quality can be true or false.

TABLE I
PREDICTION OF TEST DATA ON DATASET 2000

Actual	Prediction	
	Superior Rice Seed	Non-superior Rice Seed
Superior Rice Seed	27	3
Non-superior Rice Seed	4	26

Table 1 shows system testing results in predicting rice image test data, totaling 60 test data on a supervised learning model with data set 2000. Rice seed image prediction data will be analyzed with actual image data. The system is said to be able to provide predictions with a good level of accuracy if it can provide data predictions according to the actual data given to the system with a high percentage.

D. Accuracy Calculation

Accuracy is a measure of performance that will give the overall model's accuracy level or, in another explanation, calculate all the correct predictions from the total amount of data. Based on the results of testing the formed directed learning model, the accuracy of each classification can be calculated. The following equation can calculate accuracy calculation:

$$\text{Class Accuracy (P)} = \left(\frac{TP}{l} \right) 100\% \quad (1)$$

$$\text{Model Accuracy (P)} = \frac{\sum_{i=1}^l TP}{n} \quad (2)$$

Description:

- TP: True Positive, the number of positives the system classified as correct.
- l : Total data summary.
- n : Class count.

TABLE II
ACCURACY CALCULATION ON 2000 DATASET

Data set	Ratio	Class	Count (n)	Prediction Test Data		Accuracy (%)
				True	False	
2000	80:20	Superior	30	27	3	90%
		Non-Superior	30	26	4	86,7%
		Total				88,3%

Table 2 shows the accuracy value of the summary data that is correctly predicted by the system as calculated from the calculation of the summary data. A total of 60 image data are used as system test data. The model with a data set of 2000 and with a ratio of 80% training data: and 20% validation data has a predictive accuracy value of 88,3%.

E. Recall Calculation

A recall statistic indicates how many true positive predictions are out of the total number of true positives. Calculating the recall value must also determine the value of each class, which will then determine the average value of each recall. Recall calculation can be calculated with the following equation:

$$\text{Class Recall (R)} = \left(\frac{TP}{TP+FN} \right) 100\% \quad (3)$$

$$\text{Model Recall (R)} = \frac{\text{Total Class Recall}}{n} \quad (4)$$

Description:

- TP: True Positive, the number of positives that the system has classified as correct.
- FN: Negative data that was incorrectly classified by the system.
- n : Class count.

TABLE III
RECALL CALCULATION ON 2000 DATASET

Ratio 80:20	Superior	Non-Superior	Total
True Positive (TP)	27	26	
False Negative (FN)	3	0	
Recall (%)	90%	100%	95%

Table 3 shows the recall value of each class is calculated by calculating the number of correct and incorrect predictions. The model with a data set of 2000 and a ratio of 80% training data: 20% validation data has 90% superior rice seed recall and 100% not-superior rice seed.

F. Precision Calculation

Positive precision refers to the ratio of true predictions to overall positive predictions. To calculate the precision value, it is done by calculating the precision value of each class, then adding up and finding the average value. Precision Calculation can be calculated with the following equation:

$$\text{Class Precision (\%)} = \left(\frac{TP}{TP+FP} \right) 100\% \quad (5)$$

$$\text{Model Precision (R)} = \frac{\text{Total Class Precision}}{n} \quad (6)$$

Description:

- TP: True Positive, the number of positives that the system has classified as correct.
- FP: False Positive, positive data but incorrectly classified by the system.
- n : Class count.

TABLE IV
PRECISION CALCULATION ON 2000 DATASET

Ratio 80:20	Superior	Non-Superior	Total
True Positive (TP)	27	26	
False Positive (FP)	0	4	
Precision (%)	100%	86.7%	93.3%

Table 4 shows the recall value of each class based on the number of correct and incorrect predictions. Using 80%/20% training/validation data, the model has 100% superior precision and 86.7% non-superior precision on rice seeds. The model has a precision of 93.3%.

G. Hyperparameter Analysis

Selecting the model and tuning the hype parameters will affect the accuracy of a model in making predictions or classifying data. Dataset quality and quantity, training and validation data sharing percentages, overfitting, and underfitting determine whether a model is deployable. A system will be more accurate at pattern recognition of objects if more and better datasets are available. In addition, training data must have a greater percentage than validation data. A good model is not overfitting and underfitting. Overfitting occurs when the model predicts well on the training data but poorly on the test data. Meanwhile, underfitting if the model has a high error in the training data. In underfitting, the model cannot recognize the training data's patterns correctly.

IV. CONCLUSION

The process of determining the quality of rice seeds based on their physical characteristics of rice seeds can be carried out by applying technology based on digital image processing techniques, which consists of the classification of superior-type seeds and non-superior seeds. Machine learning with the Convolution Neural Network algorithm can classify rice seeds with a good level of accuracy by calling seed image data. This system can perform image classification consisting of 1 rice seed object. The model made by Google Collaboratory has quite a good fitting. It also has 88% for accuracy, 95% for recall, and 93% for precision. Further research needs to be done so the model can have better results.

ACKNOWLEDGMENT

This research is supported by the Ministry of Education, Culture, Research and Technology and the Indonesian Endowment Fund for Education through the Applied Scientific Research Program Funding Program.

REFERENCES

- [1] I. Chatnuntawe, K. Tantisantisom, P. Khanchaitit, T. Boonkoom, B. Bilgic, and E. Chuangsuwanich, "Rice Classification Using Spatio-Spectral Deep Convolutional Neural Network," *Comput. Vis. Pattern Recognit. Cornell Univ.*, vol. 3, no. 1, pp. 1–22, 2018, [Online]. Available: <http://arxiv.org/abs/1805.11491>.

- [2] K. Kiratiratanapruk et al., "Development of Paddy Rice Seed Classification Process using Machine Learning Techniques for Automatic Grading Machine," *J. Sensors*, vol. 2020, 2020, doi: 10.1155/2020/7041310.
- [3] S. D. Fabiyi et al., "Varietal Classification of Rice Seeds Using RGB and Hyperspectral Images," *IEEE Access*, vol. 8, no. 1, pp. 22493–22505, 2020, doi: 10.1109/ACCESS.2020.2969847.
- [4] J. Gupta and Neelam, "Identification and Classification of Rice varieties using Mahalanobis Distance by Computer Vision," *Int. J. Sci. Res. Publ.*, vol. 5, no. 5, pp. 1–6, 2015.
- [5] D. Joshi et al., "Label-free non-invasive classification of rice seeds using optical coherence tomography assisted with deep neural network," *Opt. Laser Technol.*, vol. 137, no. January, p. 106861, 2021, doi: 10.1016/j.optlastec.2020.106861.
- [6] M. M. Tin, K. L. Mon, E. P. Win, and S. S. Hlaing, "Myanmar Rice Grain Classification Using Image Processing Techniques," *Adv. Intell. Syst. Comput.*, vol. 744, pp. 324–332, 2019, doi: 10.1007/978-981-13-0869-7_36.
- [7] Z. Y. Liu, F. Cheng, Y. Bin Ying, and X. Q. Rao, "Identification of rice seed varieties using neural network," *J. Zhejiang Univ. Sci.*, vol. 6 B, no. 11, pp. 1095–1100, 2005, doi: 10.1631/jzus.2005.B1095.
- [8] Y. Ogawa, *Quality Evaluation of Rice*, Second Edi. Elsevier Inc., 2016.
- [9] P. R. Armstrong et al., "Detection of chalk in single kernels of long-grain milled rice using imaging and visible/near-infrared instruments," *Cereal Chem.*, vol. 96, no. 6, pp. 1103–1111, 2019, doi: 10.1002/cche.10220.
- [10] T. Ahmed, C. R. Rahman, and M. F. M. Abid, "Rice Grain Disease Identification Using Dual Phase Convolutional Neural Network Based System Aimed at Small Dataset," *agriRxiv*, vol. 2021, no. 1, pp. 1–14, 2021, doi: 10.31220/agrxiv.2021.00062.
- [11] P. Mookdarsanit and L. Mookdarsanit, "PhosopNet: An improved grain localization and classification by image augmentation," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 19, no. 2, pp. 479–490, 2021, doi: 10.12928/TELKOMNIKA.v19i2.18321.
- [12] Y. Toda et al., "Training instance segmentation neural network with synthetic datasets for crop seed phenotyping," *Commun. Biol.*, vol. 3, no. 1, pp. 1–12, 2020, doi: 10.1038/s42003-020-0905-5.
- [13] M. Uddin, M. A. Islam, M. Shajalal, M. A. Hossain, and M. S. I. Yousuf, "Paddy seed variety identification using T20-HOG and Haralick textural features," *Complex Intell. Syst.*, vol. 8, no. 1, pp. 657–671, 2022, doi: 10.1007/s40747-021-00545-0.
- [14] P. T. T. Hong, T. T. T. Hai, L. T. Lan, V. T. Hoang, V. Hai, and T. T. Nguyen, "Comparative Study on Vision Based Rice Seed Varieties Identification," *Proc. - 2015 IEEE Int. Conf. Knowl. Syst. Eng. KSE 2015*, pp. 377–382, 2015, doi: 10.1109/KSE.2015.46.
- [15] B. Lurstwut and C. Pornpanomchai, "Image analysis based on color, shape and texture for rice seed (*Oryza sativa* L.) germination evaluation," *Agric. Nat. Resour.*, vol. 51, no. 5, pp. 383–389, 2017, doi: 10.1016/j.anres.2017.12.002.
- [16] A. A. Aznan, R. Ruslan, I. H. Rukunudin, F. A. Azizan, and A. Y. Hashim, "Rice seed varieties identification based on extracted colour features using image processing and artificial neural network (ANN)," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 6, pp. 2220–2225, 2017, doi: 10.18517/ijaseit.7.6.2990.
- [17] H. Nguyen-Quoc and V. T. Hoang, "Rice seed image classification based on HOG descriptor with missing values imputation," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 18, no. 4, pp. 1897–1903, 2020, doi: 10.12928/telkomnika.v18i4.14069.
- [18] K. Aukkapinyo, S. Sawangwong, P. Pooyoi, and W. Kusakunniran, "Localization and Classification of Rice-grain Images Using Region Proposals-based Convolutional Neural Network," *Int. J. Autom. Comput.*, vol. 17, no. 2, pp. 233–246, 2020, doi: 10.1007/s11633-019-1207-6.
- [19] S. Kido, Y. Hirano, and N. Hashimoto, "Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN)," *2018 Int. Work. Adv. Image Technol. IWAIT 2018*, pp. 1–4, 2018, doi: 10.1109/IWAIT.2018.8369798.
- [20] Kurnianingsih et al., "Segmentation and Classification of Cervical Cells Using Deep Learning," *IEEE Access*, vol. 7, no. 1, pp. 116925–116941, 2019, doi: 10.1109/ACCESS.2019.2936017.
- [21] B. Zhao, J. Feng, X. Wu, and S. Yan, "A Survey on Deep Learning-based Fine-grained Object Classification and Semantic Segmentation," *Int. J. Autom. Comput.*, vol. 14, no. 2, pp. 119–135, 2017, doi: 10.1007/s11633-017-1053-3.
- [22] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Elsevier Comput. Electron. Agric.*, vol. 153, no. July, pp. 46–53, 2018, doi: 10.1016/j.compag.2018.08.013.
- [23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, no. 1, pp. 1800–1807, 2017, doi: 10.1109/CVPR.2017.195.
- [24] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [25] A. R. Pazoki, F. Farokhi, and Z. Pazoki, "Classification Of Rice Grain Varieties Using Two Artificial Neural Networks (MLP And Neuro-Fuzzy)," *J. Anim. Plant Sci.*, vol. 24, no. 1, pp. 336–343, 2014.
- [26] W. Liu, S. Zeng, G. Wu, H. Li, and F. Chen, "Rice seed purity identification technology using hyperspectral image with lasso logistic regression model," *Sensors*, vol. 21, no. 13, p. 4384, 2021, doi: 10.3390/s21134384.
- [27] B. Filipović, V. Panić, M. Brdar, S., & Brkljač, "Significance of Morphological Features in Rice Variety Classification Using Hyperspectral Imaging," *Int. Symp. Image Signal Process. Anal.*, vol. IEEE, pp. 171–176, 2021, doi: 10.1109/ISPA52656.2021.9552086.
- [28] S. Qadri et al., "Machine vision approach for classification of rice varieties using texture features," *Int. J. Food Prop.*, vol. 24, no. 1, pp. 1615–1630, 2021, doi: 10.1080/10942912.2021.1986523.
- [29] Y. Hamid, S. Wani, A. B. Soomro, A. A. Alwan, and Y. Gulzar, "Smart seed classification system based on MobileNetV2 architecture," in *2022 2nd International Conference on Computing and Information Technology (ICCIT)*, 2022, pp. 217–222, doi: 10.1109/ICCIT52419.2022.9711662.
- [30] N. A. Mohidem, N. Hashim, R. Shamsudin, and H. Che Man, "Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content," *Agriculture*, vol. 12, no. 6, p. 741, 2022, doi: 10.3390/agriculture12060741.