APPLICATION OF BACKPROPAGATION NEURAL NETWORK ALGORITHM FOR CIHERANG RICE IMAGE IDENTIFICATION

Dita Aprilia[1*]; Jajam Haerul Jaman[2]; Riza Ibnu Adam[3]

Informatics Engineering
[1,2,3]Universitas Singaperbangsa Karawang, Indonesia
www.unsika.ac.id
[1]dita.16075@student.unsika.ac.id, [2]jajam.haeruljaman@staff.unsika.ac.id, [3]riza.adam@staff.unsika.ac.id
(*) Corresponding Author

Abstract— Rice is a food source for carbohydrates that are most consumed in Indonesia, because of this the production is higher compared to other food crops. There are several superior rice varieties planted by the farmers, one of them is Ciherang. This type is widely planted by farmers because has high selling as economic value and can be used as premium rice. The existence of several types of rice that had a high sales value makes some person was deceitfulness by mix the rice with premium quality with bad quality. Many people do not know the problem of distinguishing types of rice from one to another that has the same shape. Classification techniques using the backpropagation neural network algorithm and image processing are used to identify one of the most preferred types of rice, Ciherang. The network architecture model on the backpropagation algorithm is very influential on the value of accuracy. In determining the best network's architectures, 4 times attempted where network architecture with 5 nodes in the input layer, 8 nodes in the hidden layer, and 1 node in output layer produce the highest accuracy of 82,66%.

Keywords: Rice Identification, Backpropagation Neural Network, Image Processing.

Abstrak— *Beras merupakan komoditas pangan sumber karbohidrat yang paling banyak dikonsumsi di Indonesia, karena hal itu produksinya lebih tinggi dibandingkan dengan tanaman pangan lainnya. Ada Beberapa varietas padi unggul yang banyak ditanam oleh petani salah satunya adalah jenis Ciherang. Banyak tengkulak yang menyukai beras jenis ini karena memiliki nilai jual yang tinggi dan dapat dijadikan beras premium. Adanya beberapa jenis beras yang memiliki nilai jual yang tinggi membuat beberapa oknum melakukan oplos beras. Banyak masyrakat yang tidak mengetahui permasalahan tersebut dikarenakan sulitnya membedakan jenis beras satu dengan lainnya sebab memiliki bentuk yang sama. Teknik klasifikasi dengan menggunakan algoritma backpropagation neural network dan pengolahan citra digunakan untuk mengidentifikasi salah satu jenis beras yang banyak disukai yaitu Ciherang. Pembuatan model arsitektur jaringan pada algoritma backpropagation neural network sangat berpengaruh terhadap nilai akurasi. Dalam penentuan arsitektur jaringan yang terbaik, dilakukan 4 kali percobaan dimana arsitektur jaringan dengan 5 node input layer, 8 node hidden layer dan 1 node output layer menghasilkan akurasi tertinggi yaitu 82,66%.*

***Kata Kunci**: Identifikasi Beras, Backpropagation Neural Network, Pengolahan Citra.*

INTRODUCTION

Rice is included in one of the staple foods for some countries in the world, particularly on the continent of Asia. In Indonesia, rice is also the most widely consumed food from other sources of carbohydrates such as corn, cassava, and sweet potato (Isnawati & Fitriyani, 2019). Rice also has many varieties of varied quality. There are some varieties of rice grown mostly by farmers such as Ciherang, Mekongga, and Inpari 33 (Handoko, 2017). Of the four previously mentioned types of rice, Ciherang is still the most popular type of rice. Ciherang is also much sought out by the middlemen for their high selling value and can be made into premium and medium rice (MC KAB INDRAMAYU, 2019).

rice distribution from farmers to consumers had a range of problems such as the mixing of rice, which was caused by the middleman and rice mills which had not done the separating of rice from the grain as it would cost more than that, a few individuals were mixing the rice good quality and rice bad quality (Handoko, 2017). This made the rice sold on the market impure or inconsistent with the original kind.

It is not recognized by the consumer because the rice has almost the same shape and makes it difficult to distinguish from one another. Therefore it was necessary to identify types of rice, one by performing a type of identification using the processing of digital image and data mining classification. Digital image processing is used in the preprocessing data and extraction features to get the data that will be used in the classification process.

Extraction features are used to get the characteristics of the rice image to be used as a dataset in the classification process. The extracted features may be geometry (Srimulyani & Musdholifah, 2019), texture (Ricardo & Gasim, 2019), and a morphological feature (Cinar & Koklu, 2019). The selection of features to be used must be adjusted to the object. In some studies, certain features can increase the accuracy of the classification algorithm (Srimulyani & Musdholifah, 2019).

The classification algorithm commonly used to identify images of an object is the backpropagation neural network (BPNN), support vector machine (SVM), and k-nearest neighbor (k-NN). BPNN algorithm can be used to identify the type of rice (Fayyazi & Monadjemi, 2017), the type of chili based on the shape of the leaves (Syaban & Harjoko, 2016), breast cancer (Mohammed et al., 2018), acute leukemia (Asadi et al., 2017), Tabacco leaves (Y. Sari & Pramunendar, 2017) and heart disease (Putra, Isnanto, Triwiyatno, & Gunawan, 2018). The BPNN algorithm has tested better than the SVM and k-NN algorithms (Singh & Chaudhury, 2016).

The combination of geometry features and morphological features can be applied together with a BPNN algorithm to identify types of rice. The study aims to review applications from the BPNN algorithm in identifying the type of Ciherang rice based on the digital image data obtained through the extraction stage features using a combination of some geometry and morphological features.

MATERIALS AND METHODS

Methodologies in the research include data acquisition data, pre-processing data, extraction features, the classification of backpropagation neural network (BPNN), and evaluation.

Data Collecting

The technique used in data collecting is the study of literature and observation. The study of literature is done by reading books, articles, or various scientific works, which deals with rice, the processing of the image particularly in the extraction of characteristics, backpropagation neural network (BPNN) algorithm, and other information. For the observation stage, it was commissioned to purchase the kind of Ciherang rice at the Koperasi Pegawai Balai Besar Penelitian Tanaman Padi (KOPKARLITAN) located on Jl. Raya 12 Sukamandi, Subang, West Java.

Data Acquisition

The stage involved the needed rice, the tools used, and the process of taking photos of rice. The rice used in the study is made up of Ciherang rice and other varieties such as IR64, Basmati, and Pandan Wangi. The selection of other kinds of rice is made at random and the shape of all the rice used is either whole or incomplete. The amount of rice taken of 250 pictures of grains, with the number of photos of Ciherang rice is 100 pictures and 150 other photos of rice are a mixture of IR64 rice, Basmati, and Pandan Wangi. Taking pictures of rice is done by putting each grain into boxes that have been coated with black cardboard. A box was assigned a hole in the top as a place for a mobile camera. The mobile camera used has a resolution of 13MP and the box used has a height of 7cm a far from the camera to object.

Preprocessing Data

These steps are done to optimize the value of extracting features at the next stage. The pre-processing has three steps like transforming the RGB image into a grayscale image, segmentation image by using thresholding methods and eliminating noise. In the process of eliminating noise, this research uses a morphological operation that's closing and opening. Operation closing and the opening is used to eliminate any noise appearing in the pictures after the thresholding process, which is marked with a white spot on the background and black spots on the object. It involves using the python programming language and the pycharm application.

Feature Extraction

Extraction features were done to get values that would be characterized by the image of Ciherang rice. The features used area combination of geometry and morphology features, which are perimeter, area, length, width, and eccentricity. Explain of each description is explained in research done by (Srimulyani & Musdholifah, 2019). Perimeter value was obtained by calculating the number of pixels that contained the edge of the object. For area value, it is obtained by calculating the number of pixels on the object or pictures as a whole over the edge. Value length and width are obtained by determining the endpoints on all four sides of rice and will be calculated by distance. The last characteristic is eccentricity, which has a value range of 0 to 1, if the object approaches a straight line, the value of eccentricity is closer to 1 and vice versa. The value of eccentricity is obtained by calculating using the semi-major axis and the semi-minor axis (1) (V. P. Sari, 2019). After the characteristics of the entire picture are obtained, the data is stored in the CSV extension file. The data used in the classification uses a backpropagation neural network algorithm. It was done with the help of library OpenCV on python programming languages.

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad \text{......} (1)$$

where:
e = value of eccentricity
a = semi major axis
b = semi minor axis

BPNN Classification

Data obtained from the extraction stage of the feature extraction will be normalized by min-max normalization (2) (Prabowo, 2018) and was labeled 1 and 0 manually where the score of 1 is Ciherang rice and 0, not Ciherang rice. Min-max normalization was done to make the data a value range of 0 to 1 because the BPNN algorithm only knows value within a range of 0 to 1.

$$s' = \frac{s - min\{s_k\}}{max\{s_k\} - min\{s_k\}} \quad \text{......} (2)$$

where:
s' = value that has been normalized
s = original value
min$\{s_k\}$ = the minimum value of data
max$\{s_k\}$= the maximum value of data

BPNN classification has two main stages of training and testing. At the training stage, the first part is to determine the input layer, the hidden layer, and the output layer. To determine the number of neurons in the hidden layer, the number of neurons in the hidden layer is a greater than the number of neurons in the input layer and less than twice the number of neurons in the input layer (3) (Hidayah, 2019), and then do as many experiments as the results obtained from these calculations.

$$l < m < 2l \quad \text{......} (3)$$

Where:
l = number of neuron in the input layer
m = number of neuron in the hidden layer

Then set parameters of input such as weights and biases that are drawn at random, maximum iteration as conditions stop, and learning rate to determine how fast networks can learn patterns. The training process is carried out to a maximum of predetermined iterations. After the training process is completed, the model that has been formed from the training process will be tested using test data.

Evaluation

The evaluation stage is done to know and measure the quality, effectiveness, and accuracy of the backpropagation neural network algorithm (BPNN) on the classification of Ciherang rice type. This research uses a confusion matrix to get accuracy (4), recall (5), precision (6), and F-measure (7). Accuracy is used to illustrate how well the model properly classifies (the correct positive and correct negative ratio) of the entire data. The recall is used to know how much rice was predictably positive of the overall positive data. The precision was used to describe the amount of correct data that has the correct ratio of the overall positive data. The F-measure is used to know how optimally the combination of recall values and precision. The conditions in the confusion matrix are shown in Table 1.

Table 1. Confusion Matrix Table

		Actual Class	
		Positive	**Negative**
Predicted Class	**Positive**	True positive (tp)	False-negative (FP)
	Negative	False-positive (fn)	True negative (TN)

Source: (Hossin & Sulaiman, 2015)

$$accuracy = \frac{tp+tn}{tp+fp+tn+fn\}} \quad \text{......} (4)$$

$$recall = \frac{tp}{tp+fn} \quad \text{......} (5)$$

$$precission = \frac{tp}{tp+fp} \quad \text{......} (6)$$

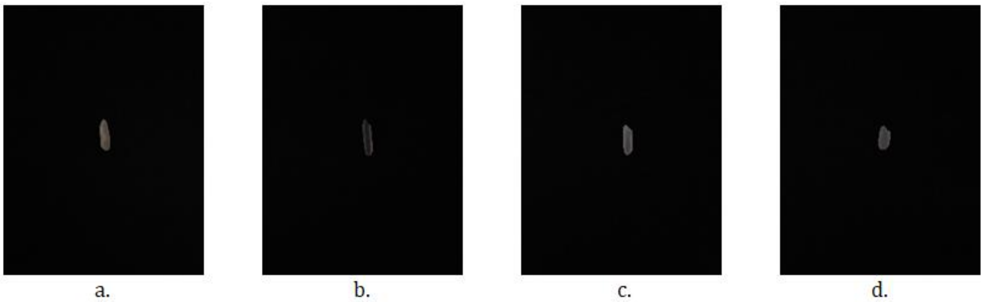
$$F - Measure = \frac{2*p*r}{p+r} \quad \text{......} (7)$$

where:
tp = the number of correct and positive predictions.
tn = the number of correct and negative predictions.
fp = the number of incorrect predictions that should be negative but predicted to be positive.
fn = the number of wrong predictions that should be positive but predicted to be negative.

RESULTS AND DISCUSSIONS

The rice used consisted of Ciherang, IR64, Basmati, and Pandan Wangi. The amount of rice in all length was 250 made up of 150 grains of

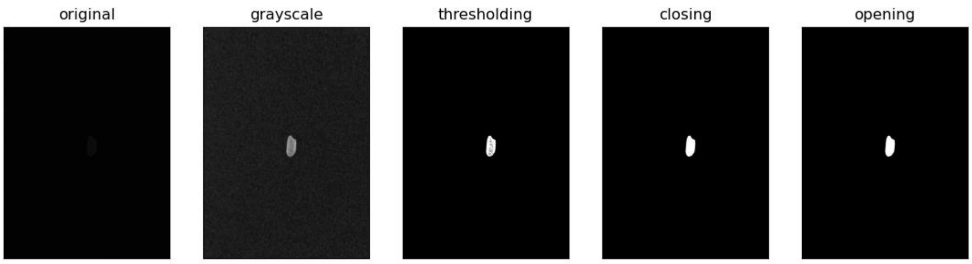
Ciherang rice and 100 grains of IR64, Basmati, and Pandan Wangi rice. Figure 1 shows a sample of each type of rice.



Source: (Aprilia, 2020)

Figure 1. a. Ciherang Rice, b. Basmati Rice, c. IR64 Rice, d. Pandan Wangi Rice

Photos of rice grains that have been obtained must be preprocessed first to get a good image because it will affect the result of feature extraction. An example of the preprocessed image is shown in Figure 2.



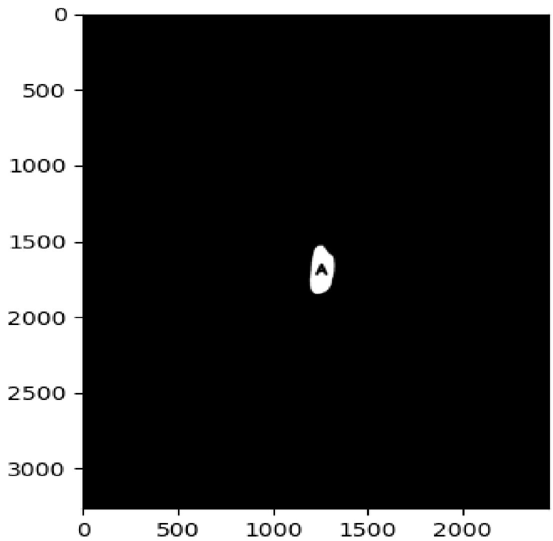
Source: (Aprilia, 2020)

Figure 2. Results of Image Processing

Figure 2 consists of the original picture which has been converted to grayscale, the one that has been enhanced using the thresholding method, and the one that has been clean from noise using a closing and opening operation.

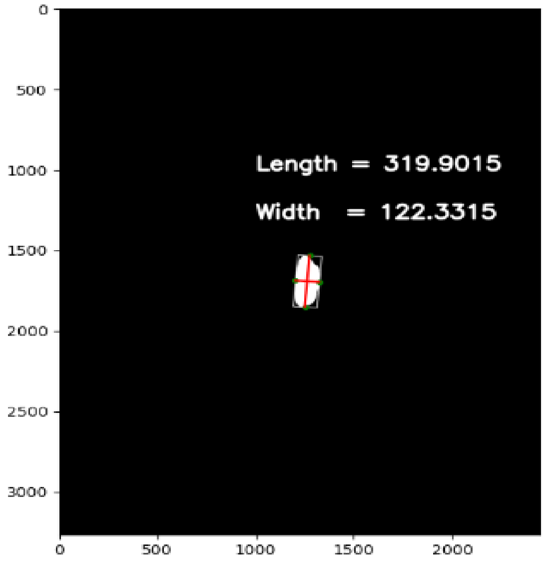
In the thresholding image, the previously grayscale image becomes black and white and shows noise on the object. Noise can be caused by such reasons as too bright or dark lighting and too transparent rice so that the black cardboard as the background is visible. In addition to the rice object, noise can also appear in the image background. For that, the closing and opening operation process is needed. There is no provision for the sequence in the use of closing and opening operation, both of which are adapted to the needs or noise in each image. In this study, the sequence that's set for all images is the first closing operation and then opening operation.

After the entire image has been through the preprocessing stage, the images are extracted to get the necessary characteristics such as perimeter, area, precision, length, width, and eccentricity. The results of each extraction feature are indicated in Figure 3, Figure 4, and Figure 5. Figure 3 shows the area, while the perimeter is the edge of the object. The red line on the object Figure 4 shows the length and width of the object is the value of length and width. In Figure 5, red lines on the object are the semi-major axis and semi-minor axis, where the value of the semi-major axis and semi-minor axis is used in eccentricity calculations (1).

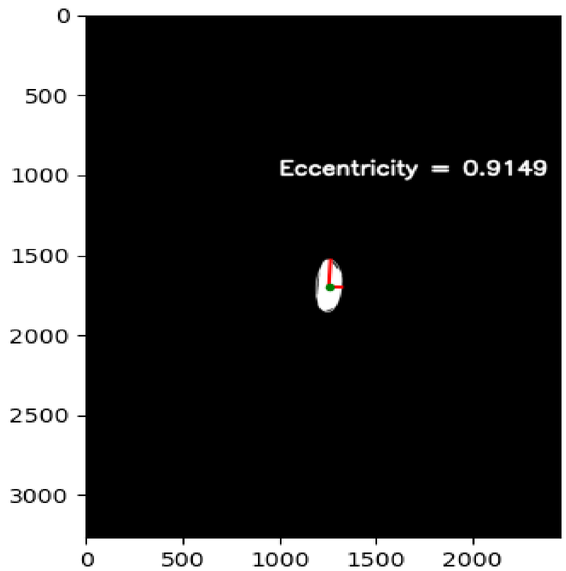


Source: (Aprilia, Jaman, & Adam, 2020)

Figure 3. Area on Object



Source: (Aprilia et al., 2020)
Figure 4. Results of the Calculation of Length and Width



Source: (Aprilia et al., 2020)
Figure 5. Results of the Calculation of Eccentricity

Data obtained from the extraction stage then normalized in a range of 0 to 1 (2). After that, the data is divided into 2, which is the training data and the test data. The division of the dataset is carried out at random using the library in the python of sklearn.model_selection with function train_test_split and use 70:30 comparisons for training data and test data in order. The training data consists of 175 rice and test data of 75 rice.

Classification is done using a BPNN algorithm. The classification process contains two main stages of training and testing. The training step is done to find the best network model by making a weight change over and over again until the condition is reached. In this study, the parameters used as input are a learning rate of 0.01 to determine how fast networks can learn the patterns on training data, the maximum iteration (max epoch) of 200 as the condition stops, and the determine the number of neurons in hidden layers (3), the study conducted 4 experiments with a different neuron in hidden layer that is 6, 7, 8, and 9. The number of neurons on the input layer is 5 by the description of the rice image and the number of neurons in the output layer is 1 because the classification is binary. In the classification process, two different activation functions are used for the sigmoid activation function and the relu activation function. The sigmoid activation function is used to get output from output layers and the ReLu activation function is used to get output from hidden layers.

Table 2. Results of the Model Training

Model	Error	Iterate	*Train* acc
[5, 6, 1]	0.330957	200	78%
[5, 7, 1]	0,320622	200	83%
[5, 8, 1]	0,196729	200	83%
[5, 9, 1]	0,453345	200	66%

Source: (Aprilia et al., 2020)

The results of the training stage can be seen in Table 2. Model columns are the architecture of a network with a different number of neurons, the first values being the neurons in the input layer, the second value or the neurons in the hidden layer, and the third value for the neuron in the output layer. The error column shows the value of error on the 200th iteration of every model made. Training shows that models with network architecture [5, 7, 1] and models with network architecture [5, 8, 1] have the same accuracy level of 83% and are the highest of the four trained models, but the error value of the models [5, 8, 1] is the smallest of all models.

At the testing stage, all models that have been through the training stage will be tested using test data that has not been used in the training process. Once the entire model has been tested, each model will be evaluated to know the network's best architectural model. An evaluation is made using a confusion matrix with value to seek is accuracy, precision, recall, and F-measure.

Table 3. Results of the Evaluation Model

Model	Accuracy	*Precision*	*Recall*	*F-Measure*
[5, 6, 1]	78,66%	71,42%	88,23%	78,94%
[5, 7, 1]	78,66%	73,68%	82,35%	77,77%
[5, 8, 1]	82,66%	83,87%	76,47%	79,99%
[5, 9, 1]	60%	75%	17,64%	28,56

Sumber: (Aprilia et al., 2020)

Table 3 shows the results of the evaluation of the four network architectural models used. The accuracy column shows the accuracy of the algorithm in identifying the type of Ciherang rice of all data in percent, and the model that has the highest accuracy is the model [5, 8, 1]. The precision column shows what percentage of the kinds of Ciherang rice is the Ciherang rice of the predicted Ciherang rice is the Ciherang rice of the predicted Ciherang rice. Recall columns show how many percent of the predicted Ciherang rice is correct from the total data of Ciherang rice. The F-measure column shows how optimally the relation between precision and recall. The model that has the highest value belongs to the model [5, 8, 1].

Table 4. Results of Accuracy Compare

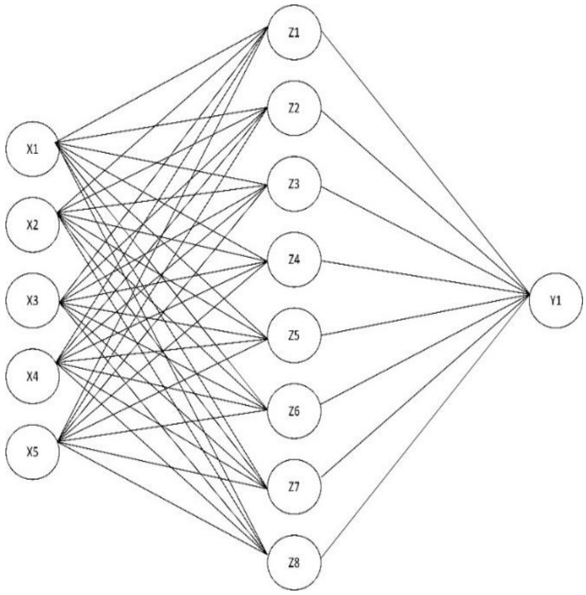
Model	Iterate	Waktu (*second*)	*Train* acc	*Test* acc
[5, 6, 1]	200	± 4,024435	78%	78,66%
[5, 7, 1]	200	± 3,889153	83%	78,66%
[5, 8, 1]	200	± 3,567708	83%	82,66%
[5, 9, 1]	200	± 3,765426	66%	60%

Source: (Aprilia et al., 2020)

After reviewing the results of each model, the next step is to compare the accuracy of the training stage and testing. The comparative chart of accuracy between the training stage and testing is shown in Table 4.

Based on the data in Table 4, the fastest execution times are owned by the model [5, 8, 1] which is ± 3,5677082 second. For accuracy between the training and testing stages, it shows that the four accuracy models at the training stage have a higher value, but the high accuracy value at the training stage does not always result in high accuracy at the testing stage. This shown in the model [5, 7, 1] that have the same accuracy with models [5, 8, 1] at the training stage which is 83%, but in the testing stages of models [5, 7, 1] results in lower accuracy than models [5, 8, 1]. At the testing stage also models [5, 7, 1] also have the same accuracy models [5, 6, 1] where models [5, 6, 1] had lower accuracy in the training stage.

As a result of the evaluation, the model with the architecture of the network's five neurons in the input layer, eight neurons in the hidden layer, and one neuron in the output layer has the best performance of the other three models. Figure 7 shows the architecture of the network of models [5, 8, 1].



Source: (Aprilia et al., 2020)
Figure 7 Network Architecture Model [5, 8, 1]

With the description of Figure 7 above as follows:

- The architecture of a network consists of one layer, a hidden layer, and an output layer.
- The input layer has five neurons made up of variables input and is represented in X1, X2, X3, X4, and X5.
- The neuron in the hidden layer is made up of eight neurons determined by conducting several experiments. Neurons in the hidden layer are symbolized by Z1, Z2, Z3, ..., Z8.
- The output layer has only one neuron containing the two labels of 1 and 0. The neurons in the output layer are represented in the Y1.

CONCLUSION

The processing of images can be used to distinguish the type of Ciherang rice by extracting features in the rice image. A backpropagation neural network algorithm can identify Ciherang rice quite well, in which four experiments conducted produced one model with the best performance that is network architecture which has five nodes in the input layer, eight-node in the hidden layer, and one node in the output layer, with a learning rate of 0.01 and a maximum iteration of 200. The evaluation of all models that have been tested has different values. The evaluation result of the best model is model [5, 8, 1] with an accuracy of 82,66%, the precision of 83,87%, recall of 76,47%, and F-measure 79,99%.

REFERENCES

Aprilia, D., Jaman, J. H., & Adam, R. I. (2020). *Laporan Akhir Penelitian: Penerapan Algoritma Backpropagation Neural Network pada Identifikasi Jenis Beras Ciherang Berdasarkan Data Citra Digital*. Karawang.

Asadi, F., Putra, F. M., Sakinatunnisa, M. I., Syafria, F., Oktafalisa, & Marzuki, I. (2017). Implementation of Backpropagation Neural Network and Blood Cells Imagery Extraction for Acute Leukemia Classification. *2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, (November), 106–110. Bandung: IEEE. https://doi.org/10.1109/ICICI-BME.2017.8537755

Cinar, I., & Koklu, M. (2019). Classification of Rice Varieties Using Artificial Intelligence Methods. *International Journal of Intelligen Systems and Applications in Engineering*, *7*(3), 188–194. https://doi.org/10.18201/ijisae.2019355381

Fayyazi, S., & Monadjemi, S. A. (2017). Identification and Classification of Three Iranian Rice Varieties in Mixed Bulks Using Image Processing and MLP Neural Network. *International Journal of Food Engineering*. https://doi.org/10.1515/ijfe-2016-0121

Handoko, D. D. (2017). Seputar Mutu Beras Kemasan dan Pencampuran Beras. Retrieved from nusakini.com website: https://www.nusakini.com/news/seputar-mutu-beras-kemasan-dan-pencampuran-beras

Hidayah, A. N. (2019). *Penerapan Jaringan Syaraf Tiruan Backpropagation Klasifikasi Penyakit ISPA (Studi Kasus : Rumah Sakit Universitas Riau)*. Universitas Islam Negeri Sultan Syarif Kasim Riau.

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *5*(2), 1–11.

Isnawati, I., & Fitriyani. (2019). *Distribusi Perdagangan Komoditas Beras Indonesia Tahun 2019* (M. Karmiati & R. Suerlianto, Eds.). Jakarta: BPS RI/ BPS-Statistics Indonesia.

MC KAB INDRAMAYU. (2019, May). InfoPublik - Ciherang Masih Jadi Jenis Padi Unggulan di Indramayu. Retrieved from InfoPublik website: http://infopublik.id/kategori/nusantara/348958/ciherang-masih-jadi-jenis-padi-unggulan-di-indramayu

Mohammed, M. A., Al-khateeb, B., Rashid, A. N., Ibrahim, D. A., Ghani, M. K. A., & Mostafa, S. A. (2018). Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers and Electrical Engineering*, *70*(August), 871–882. https://doi.org/10.1016/j.compeleceng.2018.01.033

Prabowo, T. (2018). *Implementasi Jaringan Syaraf Tiruan Backpropagation untuk Klasifikasi Serangan pada LOG Firewall (Studi Kasus : Pusat Teknologi Informasi dan Pangkalan Data (PTIPD) UIN SUSKA RIAU)*.

Putra, L. S. A., Isnanto, R. R., Triwiyatno, A., & Gunawan, V. A. (2018). Identification of Heart Disease With Iridology Using Backpropagation Neural Network. *2018 2nd Borneo International Conference on Applied Mathematics and Engineering (BICAME)*, 138–142. Balikpapan: IEEE. https://doi.org/10.1109/BICAME45512.2018.1570509882

Ricardo, D., & Gasim, G. (2019). Perbandingan Akurasi Pengenalan Jenis Beras dengan Algoritma Proagasi Balik pda Beberapa Resolusi Kamera. *RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 131–140. https://doi.org/10.29207/resti.v3i2.894

Sari, V. P. (2019). *Identifikasi Betta Fish Berdasarkan Ekstraksi Bentuk Menggunakan Parameter Eccentricity dan Metric* (Universitas Lampung). Universitas Lampung. Retrieved from http://digilib.unila.ac.id/58696/

Sari, Y., & Pramunendar, R. A. (2017). Classification Quality of Tobacco Leaves as Cigarette Raw Material Based on Artificial Neural Networks. *International Journal of Computer Trends and Technology (IJCTT)*, *50*(3), 147–150. https://doi.org/10.14445/22312803/IJCTT-V50P126

Singh, K. R., & Chaudhury, S. (2016). Efficient technique for rice grain classification using back-propagation neural network and wavelet decomposition. *The Institution of Engineering and Technology (IET)*, 1–8. https://doi.org/10.1049/iet-cvi.2015.0486

Srimulyani, W., & Musdholifah, A. (2019). Identification of Rice Variety Using Geometric Features and Neural Network. *IJCSS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(3), 301–312. https://doi.org/https://doi.org/10.22146/ijccs.48203

Syaban, K., & Harjoko, A. (2016). Klasifikasi Varietas Cabai Berdasarkan Morfologi Daun Menggunakan Backpropagation Neural Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *10*(2), 161–172. https://doi.org/10.22146/ijccs.16628