International Applied Science
UMJember Proceeding Series (2022) Vol. 1, No. 2 : 74-81

ICRD 2022

Identification of Long Bean Seed Varieties Using Digital Image Processing Coupled With Neural Network Analysis

Wahyu Nurkholis Hadi Syahputra*[1,2], Dandi Citra Nugraha[2], Abdul Jalil[3,4], Chatchawan Chaichana[1]

[1]Renewable Energy and Energy Conservation, Department of Mechanical Engineering, Chiang Mai University, Thailand
[2]Department of Agricultural Engineering, Jember University, Indonesia
[3]Department of Agronomy, University of Jember, Indonesia.
[4]Department of Agroteknology, Muhammadiyah Jember University, Indonesia.

DOI: https://doi.org/10.32528/ias.v1i2.164
*Correspondence:
Wahyu Nurkholis Hadi Syahputra
Email: wahyu@agrimart.co.id

Published: Juni, 2022

Abstract: Identification of long bean seed varieties can be used to save plant variety and intellectual property rights. Using digital image processing combined with artificial neural networks (ANN) has a possibility to recognize the seed morphology. The purpose of this research is to identify the image variables that can be used to identify long bean seed varieties so that the best algorithm of artificial neural networks can be arranged and the level of accuracy in expecting the long bean varieties. The samples used in this study were long bean seeds of parade tavi, kanton tavi, branjangan, and petiwi varieties. For each variety, 400 samples were taken for training data and 200 samples for testing data, so the total sample was 2400 long bean seeds. The research stages include image acquisition, image retrieval, image variable estimation, image processing program development, data analysis, ANN training, long bean variety identification program preparation, and program validation. The results showed that ANN with 10 hidden layers is the best model to develop a long bean seed identification. The identification program of long bean seed varieties resulting from the integration of image processing with artificial neural networks has an accuracy of 99.75%.

Keywords: *Artificial Neural Network, Computer Vision, Image Processing, Seed Variety*

INTRODUCTION

Long bean production in Indonesia from 2010 to 2016 decreased [1]. Using superior varieties in long beans is a deal to increase production and improve quality. A variety is categorized as superior if it has one or more advantages over the existing varieties in the long bean plant [2]. At this time, many foreign and domestic seed companies are interested in producing superior varieties of long beans. The manufacture of these new select varieties cannot be separated from the protection of plant varieties and intellectual property rights [3]. Indonesia makes sui generis regulations or special regulations that are effective in protecting plant varieties. An identification process is needed to maintain the genetic purity of long bean seed varieties. One commonly applied method for identifying long bean varieties is by describing the morphology of the seeds [4]. Morphology has essential characteristics in determining types such as size, color, area, and shape of seeds. Some of these characteristics take a long time, are difficult to measure manually, and must be carried out by experts. The right way to describe these characteristics is to use digital image processing combined with artificial neural networks [5], [6]. Digital image processing is an image processing technique that aims to improve image quality so that it can be easily interpreted by humans or computer machines in the form of images [7]. Image processing can provide quantitative image properties (image quality variables) needed as input for pattern recognition. Pattern recognition is the science of classifying something based on quantitative measurements of an object's main features or properties [8]. An artificial neural network is an information processing system with characteristics similar to biological neural networks [5]. Some image features used in digital image processing techniques and artificial neural networks are color, area, height, width, perimeter, and shape factor. These image features may be used as variables for identifying long bean seed varieties. The advantages of the identification process using digital image processing and artificial neural networks are that it saves time, is inexpensive, and is able to identify varieties of long bean seeds accurately.

The aims of this research are to determine the image variables that can be used as input for artificial neural networks in the identification of long bean seed varieties, to determine the best variation used as a validation process, and to determine the level of accuracy of digital image processing techniques and artificial neural networks to identify long bean seed varieties.

METHOD

Experimental Setup

This research was conducted at the precision agriculture and geoinformatics laboratory, department of agricultural engineering, Jember University, Indonesia. Supporting tools and materials are a set of image processing tools, installing IC Capture 2.2, Jasc Paint Shop Pro 9, SharpDevelop 4.4, Jupyter notebook, and long beans of the kanton tavi, parade tavi, branjangan, and pertiwi variety with a total of 600 seeds each variety.

In this research there are standardization in image capture, namely determining the distance between the camera and the object with the aim of getting the same or close to the original image, adjusting the position of the light source to the object in the best position, and determining the hue value and saturation in the IC Capture program [9], [10]. This process is carried out by determining the distance of the camera and determining the irradiation process to obtain an image of the long bean seed that is close to the original, a little shadow appears, and there is no excess light that affects the color of the object. Figure 1 shows the design of the image capture tool.

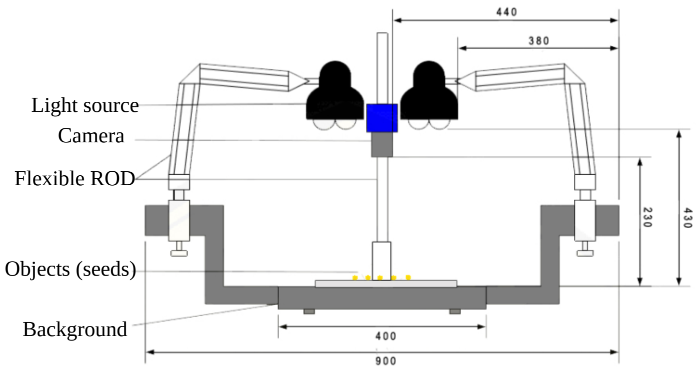
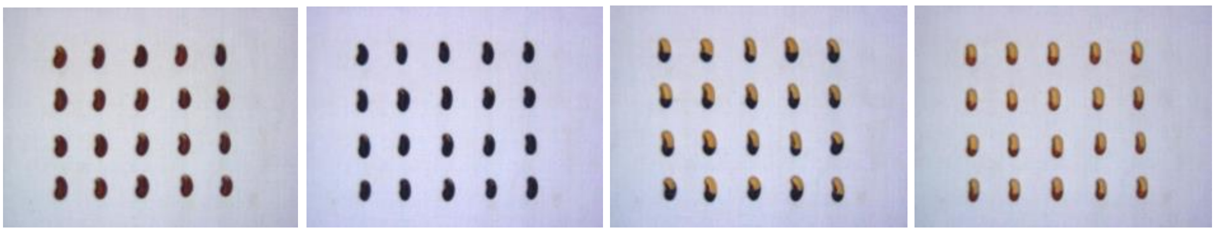


Figure 1. Experimental tools

Data Acquisition

The Image of long bean seeds was taken by using a CCD camera (charge-coupled device) in a 24-bit .bmp (bitmap) format with a resolution of 1024 x 768 pixels [11]. The bitmap format has the advantage of storing digital and complete coded image data and preserving the original data a lot. The CCD camera is set to get an image that is close to the real object. The settings made are adjusting the camera position to minimize shadow. Figure 2 shows the sample images taken from each variety of long bean seeds.



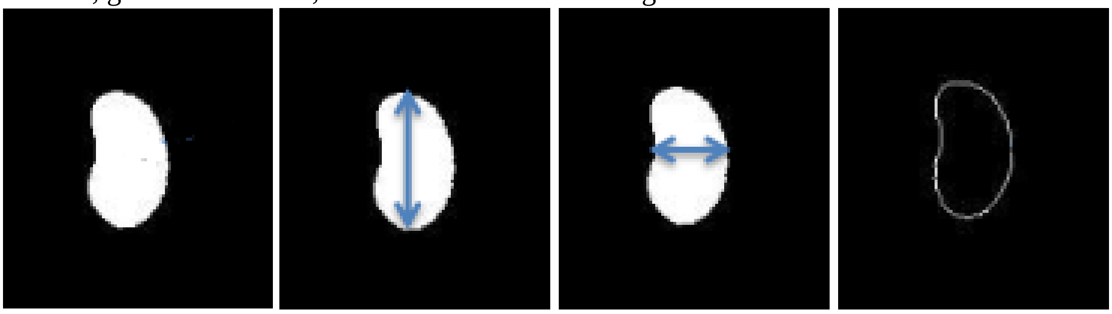
(a) (b) (c) (d)

Figure 2. Samples of long bean seeds of each variety. (a) Parade Tavi Variety; (b) Kanton Tavi Variety; (c) Branjangan Variety; (d) Pertiwi Variety

A total of 2400 long bean seed images were processed using an image processing program designed using SharpDevelop 4.4 program with the C# programming language. In one image file, there are 20 long bean seeds arranged in a 5x4 formation. The arrangement is done so that the image processing program can analyze the physical characteristics of each long bean seed with the results in the form of image variable values in one process. The image variables used as a reference to be calculated in the identification process of long bean seeds are area, perimeter, width, height, and RGB color index.

Image Processing

The image extraction process is carried out to obtain the value of the image variable that will be used for the input of the artificial neural network. The value of the image variable is obtained by performing certain operations on the pixels that make up the image. The image variables used include area, height, width, perimeter, and RGB color index. The area of the long bean seed sample is obtained from the calculation of the total number of white image pixels in the binary image. The area value is taken by assigning a limiting ordinate to each binary image of long bean seeds. The boundary ordinate is given so that the image processing program can calculate each seed area [12]. The height of the long bean seed sample was obtained from the calculation of the length of the white pixels vertically in the binary image, then calculated using the equation (*y max*- *y min*). The width of the long bean seed is obtained from the calculation of the width of the white pixels horizontally in the binary image which is then calculated by the formula (*x max*- *x min*) [13]. The perimeter value of the long bean seed is calculated from the sum of the border pixel values between the object and the background in the binary image. The perimeter value observed by giving the limiting ordinate for each binary image of long bean seeds [14]. Color index RGB was obtained from the calculation of the average value of the red color index, green color index, and blue color index in long bean seeds.



(a) (b) (c) (d)

Figure 3. Image Processing for determining (a) Area; (b) Height; (c) Width; (d) Perimeter

Data Analysis

The results of color index image extraction, area, width, height, and perimeter need to be analyzed with statistical measures to determine the correlation between physical variables and image variables [15]. Furthermore, the values of the physical variables that have been tabulated are depicted in a boxplot graph. The data sources used in the analysis of artificial neural networks are image processing data consisting of 1600 training data and 800 testing data.

The training data has a target value which is divided into 4 varieties of long beans, namely parade tavi, kanton tavi, branjangan, and pertiwi. Each data has image variables in the form of color index, area, width, height, and perimeter. Image variables that have a correlation with the physical variables of long bean varieties will be used as input for the artificial neural network. The artificial neural network training process uses the backpropagation architecture. The software used for artificial neural network analysis is a jupyter notebook [16]. The number of hidden layers used in the analysis are 10, 15, 20, and 25 hidden layers.

RESULT AND DISCUSSION

Image Variable

The results of the image variable extraction process were processed using the jupyter notebook program to obtain statistical measures of each variable. The results of the statistical measures are displayed in a boxplot graph. Through the boxplot diagram, the image variables will be known. Image variables that have an influence on differences in long bean seed varieties will be used as input for artificial neural network training. The description of the statistical size of each variety extracted from the image variable with data of 2400 samples of long bean seeds is shown in fugure 4.

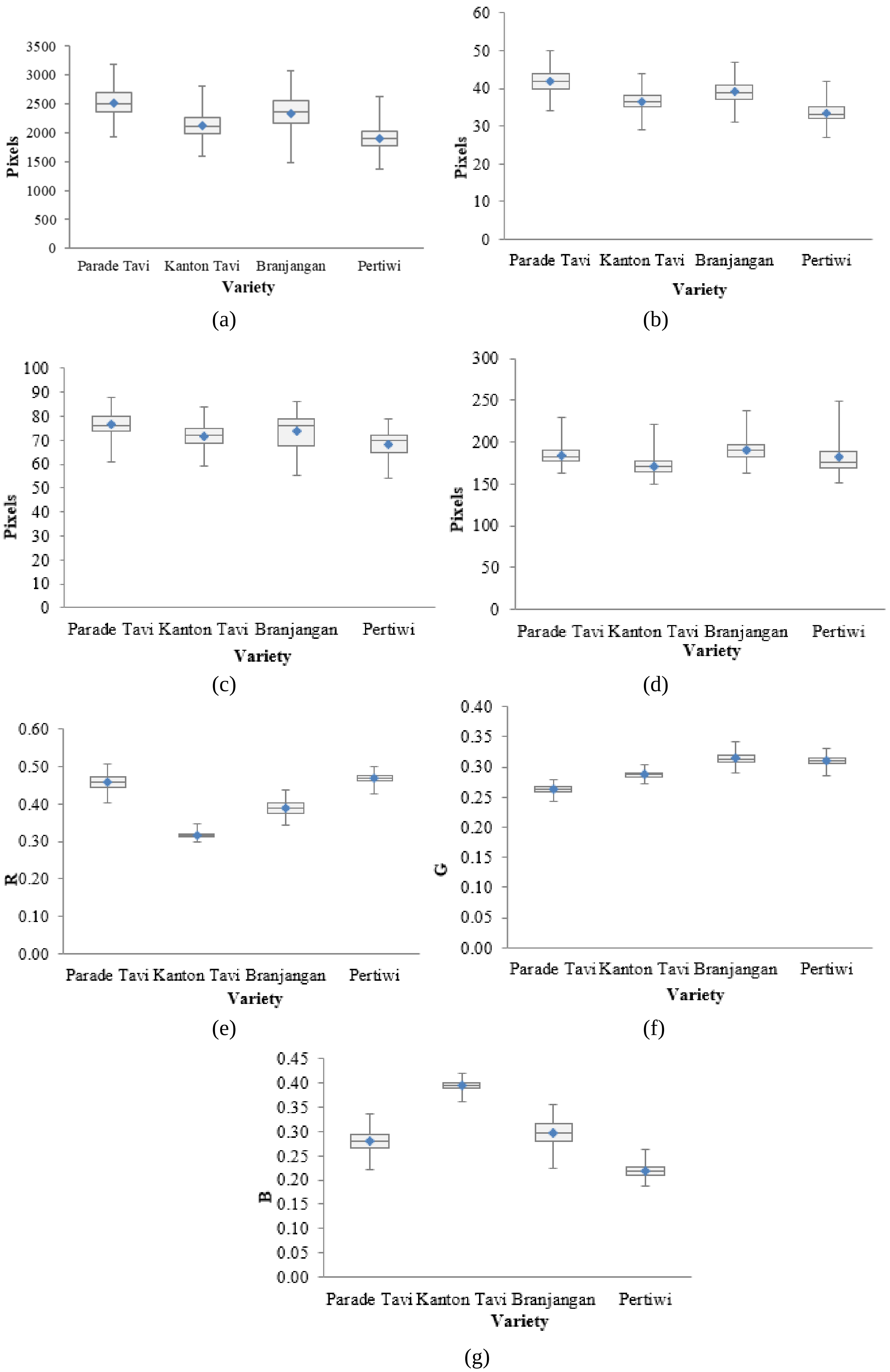


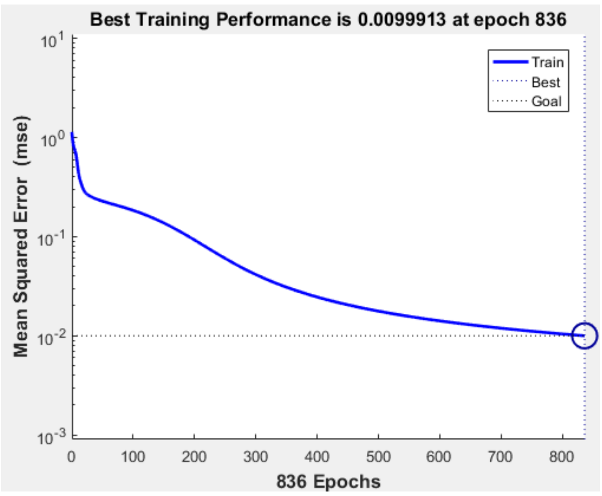
Figure 4. distribution of image variable statistical values. (a) area; (b) widht; (c) height; (d) perimeter; (e) R index; (f) G index; (g) B index.

ANN Analysis

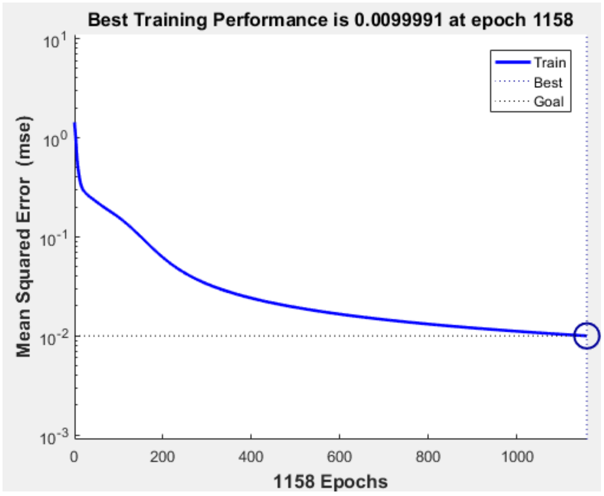
The results of statistical analysis showed that almost all of the long bean seed image variables overlapped with each other but there were still differences and could be used as input for artificial neural networks. Artificial neural networks are able to recognize patterns in the form of numbers even though the differences or changes are small. Artificial neural networks also have the ability to learn how to do work based on the data provided for training and create their own organization or representation of the information it receives during the learning time. The image variables used as input for the artificial neural network are area, height, width, perimeter, red index (R), green index (G), and blue index (B). The data sources used in the training of artificial neural networks are 1600-long bean seeds consisting of 400 seeds in each variety. Table 1 shows the results of the Mean squared error (MSE) generated at the number of different hidden layers. Figure 5 presents the ANN training result curve.

Table 1. MSE in various number hidden layer

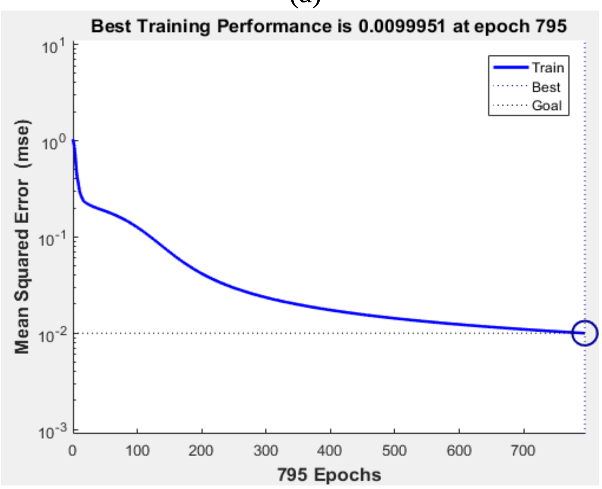
Number of Hidden Layer	Notation	MSE
10	Z1	0.0099913
15	Z2	0.0099991
20	Z3	0.0099951
25	Z4	0.0099972



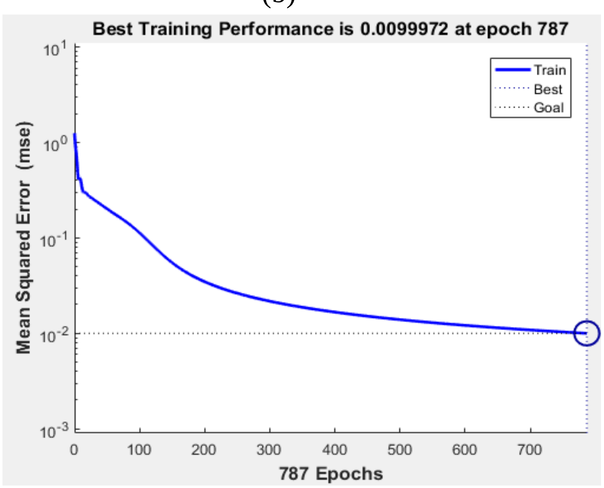
(a)



(b)



(c)



(d)

Figure 5. ANN training curve with various number hidden layers. (a) 10 hidden layers; (b) 15 hidden layers; (c) 20 hidden layers; (d) 25 hidden layers.

Selection of the best artificial neural network variation was obtained by performing a forward propagation operation on 800 samples of long bean seeds (testing data) consisting of 200 seeds for each variety. Forward propagation operations are carried out to find out how well the network is performing in target estimation [17]. The best artificial neural network variation is the variation that shows the highest suitability of target estimates. the best ANN variation is Z1 with 10 hidden layers that show the smaller MSE value, namely 0.0099913 and the accuracy rate reaches 99.63%. Table 2 presents the forward propagation data of the testing data.

Table 2. Forward propagation result

Notation	**On target each variety**				**Total**	**Accuracy (%)**
	Kanton Tavi	**Parade Tavi**	**Branjangan**	**Pertiwi**		
Z1	200	200	198	199	797	99.63
Z2	200	198	200	199	797	99.63
Z3	200	199	198	200	797	99.63
Z4	200	200	198	165	763	99.63

Validation Program

The validation process was carried out to determine the level of program accuracy in estimating long bean seed varieties. The validation process is carried out with a long bean seed variety identification program to predict seed varieties in 800 testing data. The validation of the estimation results of the long bean seed varietal identification program was analyzed using a confusion matrix table to facilitate the analysis of the program's level of accuracy. Table 3 presents the data from the ANN program validation results.

Table 3. Validation result

Variety		**Prediction**				**Total Rows**	**Accuracy prediction**	**Omission Error**
		Parade Tavi	**Kanton Tavi**	**Branjangan**	**Pertiwi**			
Actual	**Parade Tavi**	200	0	0	0	200	100	0
	Kanton Tavi	0	200	0	0	200	100	0
	Branjangan	1	0	199	0	200	99.5	0.5
	Pertiwi	0	0	1	199	200	99.5	0.5
Total Column		201	200	200	199			
User Accuracy		99.5	100	99.5	100			
Commission Error		0.5	0	0.5	0			
Total Accuracy		99.75						

The confusion matrix analysis in table 3 shows that the Parade Tavi and Kanton Tavi varieties have an accuracy rate of 100%. While the estimation of the branjangan and pertiwi varieties has an accuracy rate of 99.5%. This is in accordance with the boxplot diagram which has different variable values in each variety and is easy to distinguish even though there are some overlaps. In the color characteristics possessed by long beans, varieties of parade tavi, kanton tavi, branjangan, and pertiwi have different seed colors. These differences make it easy for the program to recognize them. The results of the validation of the long bean seed variety identification program showed a total accuracy rate of 99.75%. This means that the estimation of long bean seed varieties by the program has been maximized even though two seeds were detected incorrectly. The program which was built using the backpropagation algorithm with the variable identification of varieties has been able to distinguish each variety and correctly identify 798 seeds of long bean seeds from a total of 800 seeds.

CONCLUSION

The identification process of long bean seed varieties can be done using digital image processing with image variables of area, height, width, perimeter, color index R, G and B which are used as input for artificial neural networks. The best artificial neural network architecture for estimating corn seed varieties is 10 nodes hidden layer with activation function sigmoid bipolar, the number of iterations is 100,000, the learning rate is 0.2 and the momentum is 0.8. The validation of the artificial neural network shows that the program's level of accuracy in identifying long bean seed varieties is 99.75%.

REFERENCES

[1] I. statistic, "Luas Area, Produksi, Dan Produktivitas Kentang Di Indonesia Pada Tahun 2015-2019," *Www.Bps.Go.Id*, vol. 2019, p. 2019, 2020, [Online]. Available: https://www.pertanian.go.id/home/?show=page&act=view&id=61.

[2] K. Nooprom and Q. Santipracha, "Effect of Varieties on Growth and Yield of Yard Long Bean under Songhkla Conditions, Southern Thailand," *Mod. Appl. Sci.*, vol. 9, no. 13, p. 247, 2015, doi: 10.5539/mas.v9n13p247.

[3] A. B. Endres, "Revising Seed Purity Laws to Account for the Adventitious Presence of Genetically Modified Varieties: A First Step towards Coexistence," *J. Law Policy*, vol. 1, no. 1, pp. 131–163, 2005.

[4] A. Jordaan, D. C. J. Wessels, and H. Kruger, "Morphology, ontogeny and functional anatomy of the seeds of Colophospermum mopane," *South African J. Bot.*, vol. 67, no. 2, pp. 214–229, 2001, doi: 10.1016/S0254-6299(15)31122-4.

[5] S. Kujawa and G. Niedbała, "Artificial neural networks in agriculture," *Agric.*, vol. 11, no. 6, pp. 1–6, 2021, doi: 10.3390/agriculture11060497.

[6] Y. Sun, Z. Ren, and W. Zheng, "Research on Face Recognition Algorithm Based on Image Processing," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Mar. 2022, doi: 10.1155/2022/9224203.

[7] S. SankarNath and P. Rakshit, "A Survey of Image Processing Techniques for Emphysema Detection," *Int. J. Comput. Appl.*, vol. 114, no. 15, pp. 7–13, 2015, doi: 10.5120/20052-1983.

[8] K. Gayatri, R. D. Kanti, V. C. Sekhar Rao Rayavarapu, B. Sridhar, and V. Rama Gowri Bobbili, "Image processing and Pattern Recognition based Plant Leaf diseases Identification and Classification," *J. Phys. Conf. Ser.*, vol. 1804, no. 1, p. 012160, Feb. 2021, doi: 10.1088/1742-6596/1804/1/012160.

[9] P. G. J. Barten, "Effect of Color Saturation and Hue on Image Quality," *Soc. Imaging Sci. Technol. Image Process. Image Qual. Image Capture, Syst. Conf.*, pp. 16–21, 2003.

[10] S. Chen, D. F. Laefer, J. Byrne, and A. S. Natanzi, "The effect of angles and distance on image-based, three-dimensional reconstructions," in *Safety and Reliability – Theory and Applications*, Jun. 2017, no. June, pp. 399–399, doi: 10.1201/9781315210469-350.

[11] C. Li, X. Jia, H. Li, L. Deng, and X. Shi, "Digital image processing technology applied in level measurement and control system," *Procedia Eng.*, vol. 24, pp. 226–231, 2011, doi: 10.1016/j.proeng.2011.11.2631.

[12] C. M. Sabliov, D. Boldor, K. M. Keener, and B. E. Farkas, "Image processing method to determine surface area and volume of axi-symmetric agricultural products," *Int. J. Food Prop.*, vol. 5, no. 3, pp. 641–653, 2002, doi: 10.1081/JFP-120015498.

[13] C. Gupta, V. K. Tewari, R. Machavaram, and P. Shrivastava, "An image processing approach for measurement of chili plant height and width under field conditions," *J. Saudi Soc. Agric. Sci.*, vol. 21, no. 3, pp. 171–179, 2022, doi: 10.1016/j.jssas.2021.07.007.

[14] M. S. Gharajeh, "To Measure the Perimeter of an Ellipse Using Image Processing and Mathematical Reasoning," *Int. J. Res. Stud. Comput. Sci. Eng.*, vol. 4, no. 4, pp. 15–21, 2017, doi: 10.20431/2349-

4859.0404002.

[15] D. F. Williamson, R. A. Parker, and J. S. Kendrick, “The box plot: A simple visual method to interpret data,” *Ann. Intern. Med.*, vol. 110, no. 11, pp. 916–921, 1989, doi: 10.7326/0003-4819-110-11-916.

[16] F. Malard, L. Danner, E. Rouzies, J. G. Meyer, and E. Lescop, “EpyNN : Educational python for Neural Networks,” 2021.

[17] S. V. Razavi, M. Z. Jumaat, and A. H. EI-Shafie, “Using feed-forward back propagation (FFBP) neural networks for compressive strength prediction of lightweight concrete made with different percentage of scoria instead of sand,” *Int. J. Phys. Sci.*, vol. 6, no. 6, pp. 1325–1331, 2011, doi: 10.5897/IJPS11.204.