Mushroom Image Classification Using C4.5 Algorithm

Cucut Hariz Pratomo
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: cucutharizpratomo@gmail.com

Widyastuti Andriyani
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: widya@akakom.ac.id

This study applied five types of Mushrooms, they are Button mushrooms, Wood Ear mushrooms, Straw mushrooms, Reishi mushrooms and Red Oyster mushrooms. The feature extraction used is Order 1 with the parameters of mean, skewness, variance, kurtosis, and entropy. The process carried out to identify mushroom images by preparing image objects. There were 15 images of each mushroom class were taken for each mushroom and stored in .jpg format. The image processing is carried out by a feature extraction process. Then five images for each mushroom class are chosen. They were used as test images which will be classified so that identification results are obtained. This study applies the Classification Algorithm C4.5 to build a decision tree, which will also identify the results of the accuracy of processed mushroom images. The obtained result of accuracy was 84% in the classification of feature extraction Order 1.

***KeyWords**: C4.5, Image, Extraction Features, Mushroom, Classification, Order1*

This Article was:
submited: 20-02-23
accepted: 20-04-23
publish on: 20-07-23

How to Cite:
Pratomo, C.H., et al, "Mushroom Image Classification Using C4.5 Algorithm", Journal of Intelligent Software Systems, Vol.2, No.1, 2023, pp.17-19, 10.26798/jiss.v2i1.123

1 Introduction

Mushrooms are plants which do not have chlorophyll so they are called heterotrophy which digest their food outside the body. As living things, heterotrophic fungi can be obligate parasites, facultative parasites, or saprophytes. There are many types of mushrooms with each shape, feature, texture and color. Research on mushrooms using feature extraction and classification to identify the type of mushroom has been carried out; five images of mushroom types were identified to determine the type, including Button mushrooms, Wood Ear mushrooms, Straw mushrooms, Reishi mushrooms and Red Oyster mushrooms. The data collection was carried out by doing observation in Jejamuran Sleman Yogyakarta.

In general, Button mushrooms have white, moon-shaped and fine-textured physical characteristics. Wood Ear mushrooms have a characteristic of dark brown color, are shaped like ears and have a soft texture. Straw mushrooms have a characteristic of black color, are round in shape and rather rough in texture. Reishi Mushrooms have a characteristic of ivory yellow color, shaped like a shell and have hard texture. Red Oyster mushroom has a characteristic of red color, is wave-shaped like a flower, and has a soft and fibrous texture. The examples of mushroom images applied in this study are shown in picture 1 Button Mushroom, picture 2 Wood Ear Mushroom, picture 3 Straw Mushroom, picture 4 Reishi Mushroom, and picture 5 Red Oyster Mushroom[1][2].

Fig. 1 Button Mushroom (Jamur Kancing)

Fig. 2 Wood Ear Mushroom (Jamur Kuping)

Fig. 3 Straw Mushroom (Jamur Merang)

Fig. 4 Reishi Mushroom (Jamur Lingzhi Kerang)

Fig. 5 Red Oyster Mushroom (Jamur Tiram Merah)

In Figure 1 is a Button Mushroom with a general characteristic of being brightly colour and finely textured. Mushroom which has Latin name Agaricus bisporus is a mushroom that is edible. The wood ear mushroom in Figure 2 has a general characteristic of being dark in color and having a soft and chewy texture when it is

fresh. Mushrooms which has Latin name as Auricularia auricula are also safe to eat. In Figure 3 Mushrooms or Volvariella volvacea have the general characteristic of being dark in colour and a bit rough in texture, have fairly high protein content and are quite popular edible mushrooms. In Figure 4 the Reishi Mushroom has the general characteristics of being brightly colour and hard-textured and shaped like a clam shell. Mushroom which has Latin name Ganoderma lucidum is included in consumption mushrooms but tends to be used as traditional medicine by processing the nutritional essence inside. Moreover, the Red Oyster Mushroom in Figure 5 has the general characteristic of being brightly colored, soft and fibrous in texture, has the Latin name Pleurotus ostreatus.

2 Methodology

2.1 Order 1. The first-order feature extraction is a feature-taking method based on the characteristics of an Image Histogram, in which an Image Histogram illustrates the probability of the appearance of a pixel's gray degree value in an image[3]. The values generated on the Histogran can be calculated for several parameters including the mean, skewness, variance, kurtosis, and entropy[4].

a. Mean (μ)
Indicates the size of the dispersion of an image.

$$\mu = \sum_{n} fnP(fn) \quad (1)$$

In which fn is a gray intensity value, while p(fn) denotes a histogram value (the probability of the intensity appearing in the image).

b. Variance (σ^2)

$$\sigma^2 = \sum_{n} (fn - \mu)^2 P(fn) \quad (2)$$

Shows the elements variation in the histogram of an image.

c. Skewness (α^3)
Illustrates the relative sloping level of the histogram curve of an image.

$$\alpha^3 = \frac{1}{\alpha^3} \sum_{n} (fn - \mu)^3 P(fn) \quad (3)$$

d. Kurtosis (α^4)
Indicates the level of relative sharpness of the histogram curve of an image.

$$\alpha^4 = \frac{1}{\alpha^4} \sum_{n} (fn - \mu)^4 P(fn) - 3 \quad (4)$$

e. Histogram (H)
Indicates the size of the irregular shape of an image.

$$H = -\sum_{n} P(fn).^2 log P(fn) \quad (5)$$

2.2 C4.5 Classification. . C4.5 is a Decision Tree algorithm, ID3 evolution, presented by the same developer, using Information Gain and Gain Ratio as separation criteria (Rokach and Maimon, 2014; Rani, 2022). The Separation end when the instances number is split is below a certain threshold. The stages in the C4.5 algorithm for constructing a decision tree are as follows:

a) Select attribute as root
b) Create the branch for each value
c) Divide cases into branches
d) Repeat the process for each branch until all cases on the branch have the same class. To select an attribute as the root, the highest gain value must be found among the existing attributes, with the formula:

$$Gain(S, A) = Entrophy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} .Entrophy(S_i) \quad (6)$$

Description:
S : Set Case
A : Attribute
S_i : Number of cases on the i-th partition
n : sum partition of attribute A
|S| : number of cases in S

3 Analysis and Discussion

The implementation process consists of two stages, as follows data pre-processing and classification using the C4.5 Algorithm. In the initial process, the mushroom image data will be converted from a color data format to a grayscale image format[5]. Then from the feature extraction process, for Order 1 features, values are obtained as in Table 1 and entropy calculations produce values as in Table 2.

Table 1 Order 1 Extraction Features

Data	Order 1				
	M	V	S	K	E
Kancing1.jpg	5.8215	4.6837	0.3686	1.6222	2.6164
Kancing2.jpg	5.7515	5.2284	0.3374	1.5393	2.6159
Kancing3.jpg	5.6938	4.3638	0.5584	1.89602	2.506005
Kancing4.jpg	5.6389	4.397	0.6188	1.9117	2.4476
Kancing5.jpg	5.3643	5.4149	0.6039	1.8025	2.4954
Kuping1.jpg	6.04715	9.1906	-0.1255	1.1443	1.9083
Kuping2.jpg	4.9378	7.455	0.6842	1.7193	2.1271
Kuping3.jpg	5.1982	9.0985	0.3724	1.3046	2.0917
Kuping4.jpg	5.71805	10.331	-0.0187	1.1105	1.8981
Kuping5.jpg	4.99109	8.8146	0.5221	1.4317	2.0278
Linghi1.jpg	4.5756	6.7426	0.8901	2.1739	2.3217
Linghi2.jpg	5.38706	7.7204	0.3445	1.3965	2.3642
Linghi3.jpg	5.1344	6.76409	0.5491	1.6965	2.5029
Linghi4.jpg	4.8182	6.0659	0.7655	2.1285	2.5438
Linghi5.jpg	4.7575	6.8473	0.8076	1.9822	2.27007
Merang1.jpg	4.3124	6.6117	1.1675	2.5899	1.8298
Merang2.jpg	3.8657	7.1702	1.3093	2.8941	1.6848
Merang3.jpg	4.1401	8.0947	1.0313	2.2351	1.8588
Merang4.jpg	4.1183	7.8812	1.0843	2.328	1.7569
Merang5.jpg	4.3256	7.574	1.0098	2.2205	1.9161
Tiram1.jpg	7.6223	4.0468	-1.1147	2.7561	1.9658
Tiram2.jpg	7.553	3.9504	-0.9833	2.4407	2.0317
Tiram3.jpg	6.2239	5.0091	0.0171	1.624	2.5863
Tiram4.jpg	6.3953	6.3024	-0.2512	1.5063	2.5023
Tiram5.jpg	5.8943	6.0937	0.1074	1.4501	2.6371

Table 2 Entrophy Class for Order 1

Mushroom Class	Result
Jamur Kancing	0,795040279
Jamur Kuping	0,721928095
Jamur Lingzhi Kerang	0,760167503
Jamur Merang	0,721928095
Jamur Tiram Merah	0,721928095

Table 3 Confusion Matrix Classification Feature Extraction Order 1

No	Actual Testing Data	Predictions Jamur Kancing	Jamur Kuping	Jamur Lingzhi	Jamur Merang	Jamur Tiram Merah
1	Kancing	1	0	2	0	2
2	Kuping	0	5	0	0	0
3	Lingzhi	0	0	5	0	0
4	Merang	0	0	0	5	0
5	Tiram Merah	0	0	0	0	5

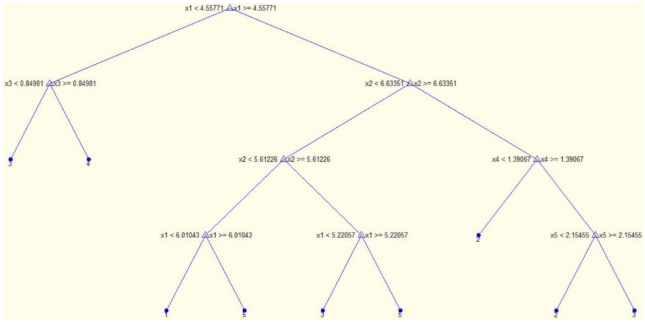

Fig. 6 Classification Tree of Feature Extraction Order 1

4 Conclusions

The identification results of Button Mushrooms were detected correctly as one mushroom image as Button mushroom. Four mushroom images were detected wrong, they were two mushroom images as Reishi mushroom, and two mushroom images were detected as Red Oyster mushroom. The results of the identification of Wood Ear Mushroom were correctly detected by five images of mushrooms as Wood Ear mushroom. The results of the identification of the Reishi Mushroom detected correctly those five images of the mushroom as the Reishi Mushroom. The results of identification of Straw Mushrooms were correctly detected by five images of mushrooms as Straw Mushrooms. The identification results of the Red Oyster Mushroom were correctly detected by five images of the mushroom as Red Oyster Mushroom.
The form of the classification tree result will appear in Figure 6. Subsequently the accuracy results are as follows: (21+0)/25 * 100% = 84% and precision values: (21 / (21+0)) * 100% = 100% and has a recall value: (21 / (21+0)) * 100% = 100%.
The accuracy value obtained is moderate, whereas, the value of precision and recall are great because there is no data as FP to look for precision and also no FN data to determine the recall value[9].

References

[1] Cong, P., Feng, H., Lv, K., Zhou, J., and Li, S., 2023, "MYOLO: A Lightweight Fresh Shiitake Mushroom Detection Model Based on YOLOv3," doi: 10.3390/agriculture13020392.

[2] Gustina, S., Fadlil, A., and Umar, R., 2017, "Sistem Identifikasi Jamur Menggunakan Metode Ekstraksi Ciri Statistik Orde 1 dan Klasifikasi Jarak," Techno.Com, **16**(4), pp. 378–386.

[3] Nixon, M. S. and Aguado, A. S., 2012, *Feature Extraction and Image Processing for Computer Vision*.

[4] Nurhayati, O. D. and Windasari, I. P., 2016, "Stroke Identification System on the Mobile Based CT Scan Image," *ICITACEE 2015 - 2nd International Conference on Information Technology, Computer, and Electrical Engineering: Green Technology Strengthening in Information Technology, Electrical and Computer Engineering Implementation, Proceedings*, Computer, and Electrical Engineering: Green Technology Strengthening in Information Technology, Electrical and Computer Engineering Implementation, Proceedings, pp. 113–116. Available at, pp. 113–116, doi: 10.1109/ICITACEE.2015.7437781.

[5] Fadlil, A., 2012, "Sistem Pengenalan Citra Jenis-Jenis Tekstil," Spektrum Industri: Jurnal Ilmiah Pengetahuan dan Penerapan Teknik Industri, **10**(1), pp. 19–29.

[6] Lee, C. H., Choi, D., Pecchia, J., He, L., and Heinemann, P., 2019, *Development of a mushroom harvesting assistance system using computer vision*.

[7] Rani, U. D., 2022, *Mushroom Classification: A Comparison of Classification Algorithms using Machine Learning Techniques*, Vol. 9, www.jetir.org.

[8] Maimon, O. Z. and Rokach, L., 2015, *Data Mining with Decision Trees: Theory and Applications (Second Edition) (Machine Perception and Artificial Intelligence)*, WSPC.

[9] Sofia, V., 2011, "Confusion Matrix-based Feature Selection Sofia Visa," ConfusionMatrix-based Feature Selection Sofia, **710**(January), p. 8.