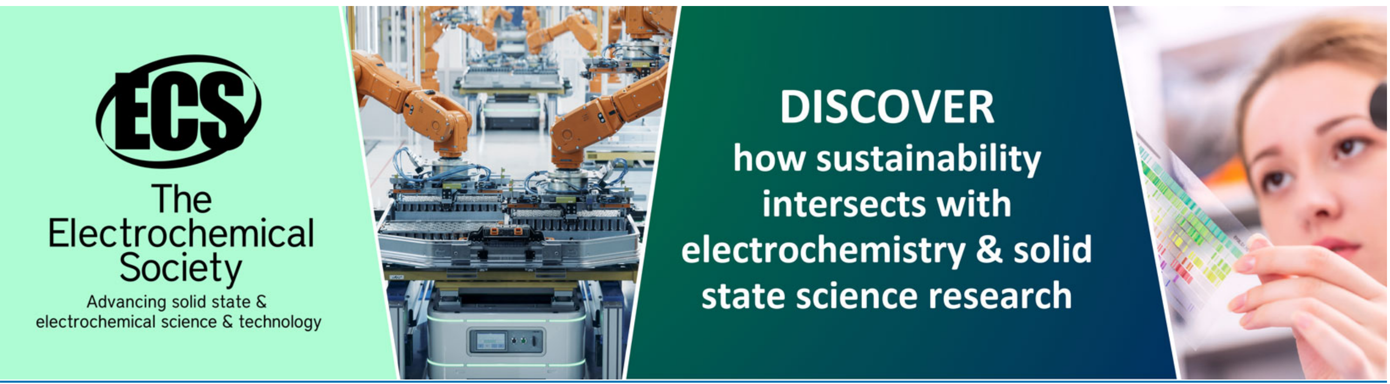
IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

Fungus image identification using K-Nearest Neighbor

To cite this article: R F Rahmat *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **420** 012097



IOP Conf. Series: Materials Science and Engineering **420** (2018) 012097 doi:10.1088/1757-899X/420/1/012097

Fungus image identification using K-Nearest Neighbor

R F Rahmat[1*], T Aruan[1], S Purnamawati[1], S Faza[1], T Z Lini[1], Onrizal[2]

[1]Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
[2]Faculty of Forestry, Universitas Sumatera Utara, Medan, Indonesia

*E-mail: romi.fadillah@usu.ac.id

Abstract. A fungus is a type of plants that have no chlorophyll, and its life depends on other living things. The fungus has various types. Some mushrooms are useful and utilized in the field of culinary, health, economy and agriculture, but there are poisonous ones that can cause infection and even death to humans. Research on fungus identification is insufficient, and the conventional method still becomes the main choice in identifying them. Therefore, an alternative is required to identify the poisonous mushrooms. The method proposed in this study is K-Nearest Neighbor. Mushroom image served as the input image of this image processing process. Before identification process, the image will go through image pre-processing step using gray-scaling, image segmentation using canny edge and thresholding, and feature extraction process using zoning method. 40 images of fungus plants were used to test the system. The accuracy rate of the system in poisonous mushroom identification is 90%.

1. Introduction

Mushroom is a plant with no chlorophyll, and its life depends on other living things. This plant also acts as the decomposer of organism food environment. A fungus can also be consumed by the human as a food supplement because of its excellent taste and good nutritional value [1]. There are various types of mushrooms scattered throughout the world, among them causing disease in humans and plants, even containing toxins [2]. There are more than 1,500,000 mushroom species in the world, and only about 74,000 of them are identified. Some fungi that have been identified are beneficial in the field of food, health, economy, agriculture, while others can cause infections to humans [3].

A fungus has specific characteristics in size, color, the shape of the hood and stem that are essential features for identification [4]. However, there is still some shortages in mushroom identification due to the various types of the fungus itself, limited information on fungus knowledge, and lack of mushroom experts. Previous research of mushroom plants identification system was built on a mobile device using the Support Vector Machine (SVM) method and several feature extractions such as histogram features and shape features [5]. Further study was conducted on fungi growth identification using Learning Vector Quantization (LVQ) method [6]. Research on edible fungus identification was performed using Principal Component Analysis (PCA) method, and it generated an accuracy rate of 85% -96% [7].

Numerous methods can be applied for identification process, one of which is K-Nearest Neighbor (KNN) to classify the stock index movements of BSE-SENSEX and NSE-NIFTY from January 2006 to May 2011 [8].

Further study was conducted in students grade input process with optical character

IOP Conf. Series: Materials Science and Engineering **420** (2018) 012097 doi:10.1088/1757-899X/420/1/012097

recognition (OCR) using K-Nearest Neighbor. The testing data in this research were cohered to training data through the nearest distance. A Research on heart disease also implemented K-Nearest Neighbor method. This classification used seven types of data sets of Andhra Pradesh community data [9].

Based on the previous research, authors proposed an alternative to identify poisonous fungus using K-Nearest Neighbor for identification process and Zoning for feature extraction. This method is expected to enhance the performance of the identification process of fungi species and to boost its accuracy rate for the identification of poisonous fungus becomes more efficient.

Mushrooms are very diverse and useful in the field of food, health, economics, and agriculture. However, some of them can cause infection and lead to human deaths. The knowledge on identification of these plants is still insufficient due to the plant variations, limited information on fungus characteristics, and lack of mushroom experts available. Conventional way still becomes the main option in mushrooms identification. Therefore, an approach is needed to identify the poisonous fungus plants.

Various researches on fungus identification has been conducted in the past. One of the studies is the implementation of Support Vector Machine for mobile-based mushrooms detection using Histogram and Principal Component Analysis (PCA) as extraction feature methods [5]. Feature extraction process in this study used 3-3-2 bit RGB histogram palette; the mushroom hood section is divided into five sections from left to right as well as parts of the mushroom bodies. The research used 35000 images of mushrooms plants.

2. Methodology

The identification of poisonous mushroom species in this study consists of several stages. The stage begins with the image acquisition process of toxic and non-toxic fungus images to be used as the training and testing data. The next step is the preprocessing stage consisting of image resizing; image gray-scaling; image segmentation using threshold and canny edge detection; feature extraction using Zoning method; and identification process using the K-Nearest Neighbor method.

After all the steps were completed, the system will generate an output in the form of information about the type of fungus along with information whether a poison contained in the mushroom or not. The general architecture of this system can be seen in figure 1.

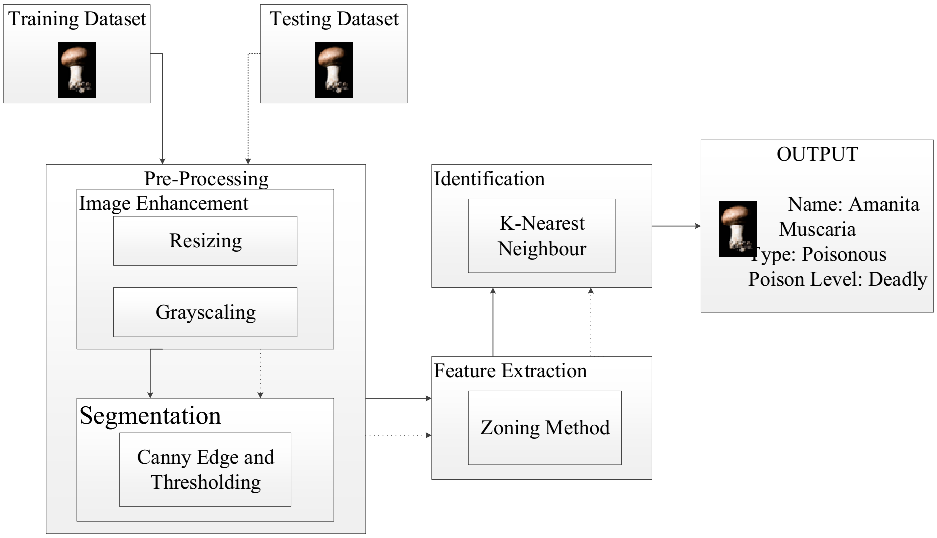


Figure 1. General architecture

2.1. *Image acquisition and Preprocessing*

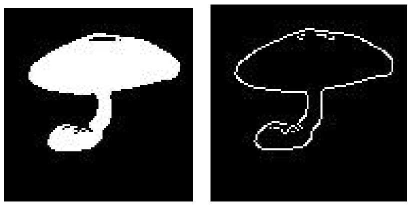
Image acquisition is a process to acquire mushroom images for the input of the system. The images were taken using a camera with good resolution then stored in the desktop (see Figure 2a). The images are in .jpg and .jpeg format. After image acquisition, the images will be normalized and improved the quality through image enhancement and image segmentation process. Image resizing was performed to obtain focus area of the mushroom images by manipulating the object orientation. The image will be resized to 100x100 pixels (see Figure 2b). Gray scaling was conducted for the image to be processed in the next stage of feature extraction by converting the RGB image of a fungus into a gray image (see Figure 2c) where it was given a range of threshold values to limit the gray level.

(a) (b) (c)

Figure 2. (a) Original image; (b) Resized image (c) Greyscale Image

2.2. *Image segmentation*

After preprocessing, the images will go through image segmentation process using canny edge detection and threshold methods to separate object from its background so it will be easier to analyze the images [11]. This stage aims to obtain the bit value of an image by separating the mushroom object from the background. Canny edge detection has the purpose of emphasizing the boundary lines of an object in the image. Edge detection in this research is a result of the changes in gray degrees. The result of the thresholding process and edge detection can be seen in Figure 3.



(a) (b)

Figure 3. (a) Thresholding result, (b) Canny Edge Detection

2.3. Feature extraction

Feature extraction using Zoning method was performed to obtain the vector of an image. Feature extraction was performed using zoning method since the image of canny edge detection is divided into regional blocks so that for each region obtained a feature that describes its characteristics. The steps of this method are as follows:

- Calculate centroid value of input image that expressed by 1 and 0.
- Divide the input image into the same n zone. Initial image size of 100x100 pixels formed a 100x100 matrix with zone partition of n=10x10.
- Input bit value of the image into the matrix. For example, the input is an image with a size of 10 x 10 matrix for each row, then matrix dimension of n = 3 x 3 was used in the zone partition. For the remaining matrix, added a matrix multiplication as required, in the picture below, it takes two additional matrix row to do the calculation. So the final matrix size of the image is 12x12 (see Figure 4a).

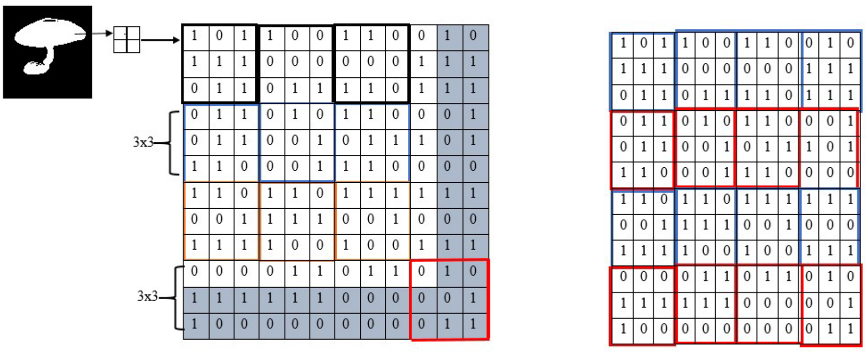


Figure 4. (a)12x12 matrix multiplication (b) Fixed matrix zone

- Repeat b and c to all existing pixels in the zone as in Figure 4(b) until zone n=3x3 is fulfilled.
- Calculate the frequent feature values found (1 or 0) in each 10x10 matrix zone. If the ratio of the feature value of 1 and 0 is 1:1 then specify the value to 1. Zoning result can be seen in Figure 5.

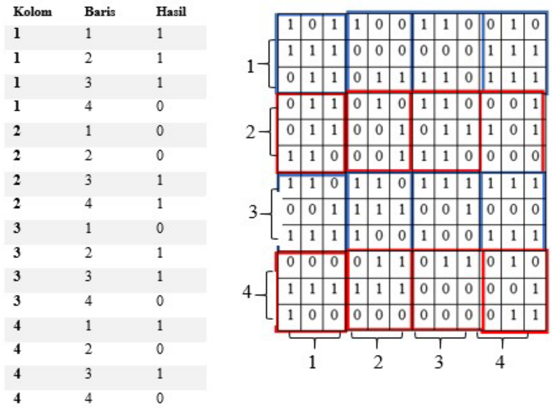


Figure 5. Zoning result

2.4. *K-Nearest Neighbour Implementation*

The value obtained from the extraction feature was stored to be used for system testing by comparing both values of training and testing images. K-Nearest Neighbor method was used to identify whether the mushroom is poisonous or edible. The KNN step is divided into two parts, i.e., training and testing. Data training is a data validation process. 40% of total data were used as training data. Identification process using testing data was performed after training data were validated. The calculation of identification process is as follows:

- Set parameter k = amount of nearest neighbors
- Input the first testing data
- Calculate the distance between new data and training data.
- Sort the distance and determine the nearest neighbor based on the minimum distance of- K.
- Determine the category of the nearest neighbor value
- Use the majority category of the nearest neighbor value as the predicted value of the new data.

In the system testing, the input image has already went through zoning process in previous feature extraction phase with feature values of 1,1,1,0,0,0,1,1,0,1,1,0,1,0,1,0 .

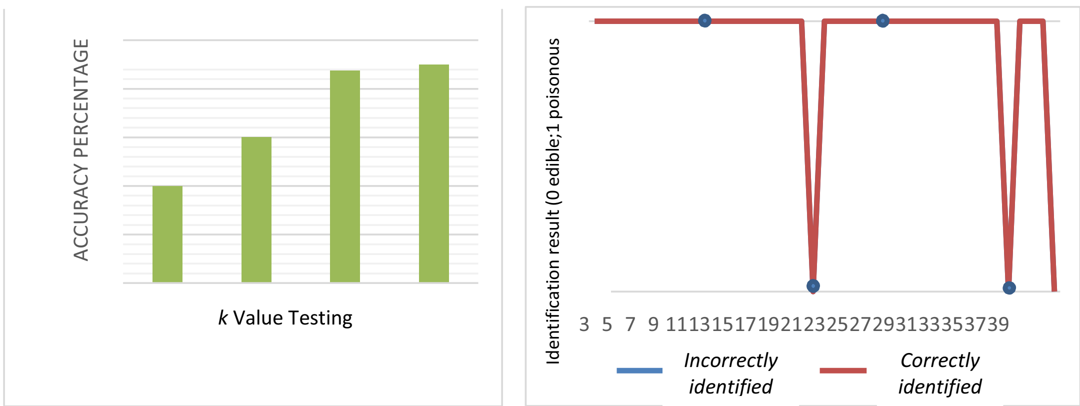
3. Result and analysis

The data used in this research are mushroom images acquired from mushroom expert and mushroomserver.org with eight types of mushroom; namely agaricus bisporus, amanita phalloides, amanita muscaria, cortinarius rubripes, floccularia fusca, gyromitra esculenta, galerina marginata, volvariella speciosa. They were divided into two categories of poisonous and edible fungus. The data summary of this research can be seen in table 1.

Table 1. Result of fungus testing

Image Name	Note		Toxic Level
	Poisonous	Edible	
mshrm_srvg_0x_ab_(1).jpg		√	-
mshrm_srvg_0x_ab_(2).jpg		√	-
mshrm_srvg_0x_am_(1).jpg	√		2-3 hours. Causes swelling of the liver and cessation of bile.
mshrm_srvg_0x_am_(2).jpg	√		2-3 hours. Causes swelling of the liver

		and cessation of bile
mshrm_srvg_0x_ge_(1).jpg	√	6-12 hours. Causes damage to the digestive organs.
mshrm_srvg_0x_ge_(2).jpg	√	6-12 hours. Causes damage to the digestive organs.
mshrm_srvg_0x_gm_(1).jpg	√	30 minutes – 2 hours. Damage the respiratory system, brain, and kidney.
mshrm_srvg_0x_gm_(2).jpg	√	30 minutes – 2 hours. Damage the respiratory system, brain, and kidney.
mshrm_srvg_0x_cr_(1).jpg	√	4-15 days. Causes kidney malfunctioning.
mshrm_srvg_0x_cr_(2).jpg	√	4-15 days. Causes kidney malfunctioning.



(a) (b)

Figure 6. (a) *k* value testing (b) graph of identification result

We tried to test all 40 data, which shows significant result, based on the test performed in mushroom identification using K-Nearest Neighbor, we obtained an accuracy result of 90%, where 36 identified as right classification from 40 data.

A test was conducted using various k values, k = 1, k = 3, k = 4, k = 5 as shown in Figure 6a. The result of each test is different; it is found that the higher the value of k, the more accurate the system, whereas more training data to be processed. The amount of images used for testing data is 5 for each type with a total of 40 images. Figure 6b shows errors in data 11, 20, 26, and 37.

In data 11, it shows that volvariella speciosa is a poisonous mushroom where it supposed to be a non-poisonous one. Another error generated in data 20 and 26. The system considered the gyromitra esculenta and galerina_marginata as edible mushrooms while they should be classified as poisonous mushrooms. The system could not identify data 37 of cortinarius rubripes. These errors may cause the lack of image sharpness (data 11), the similarity of features in some images (data 37), as well as the poorly trained data (data 11, 20, 26 and 37).

IOP Conf. Series: Materials Science and Engineering **420** (2018) 012097 doi:10.1088/1757-899X/420/1/012097

4. Conclusion

Based on the research on poisonous mushroom identification using K-Nearest Neighbor, it can be concluded that K-Nearest Neighbor can identify the poisonous mushrooms with an accuracy rate of 90%. The parameters used to support the method are 40 images of 8 different types of mushroom, k value with the nearest minimum distance. The higher the k value, the higher the accuracy rate obtained.

References

[1] Hasanuddin 2014 Macroscopic Wood Fungus For Biology Learning Media *Biotic Journal* 2(**1**) pp 38-52

[2] Parjimo H and Andoko A 2007 *Fungi Cultivation (Ear Mushroom, Oyster Mushroom, Merang Mushroom)* (Jakarta: AgroMedia)

[3] Tyasrini E, Triswaty W, and Susantina 2006 Relationship between Properties and Metabolites Candida spp with Pathogenesis of Candidiasis *Maranatha Medical Journal* 6(**1**) pp 52-66.

[4] Smith A H and Weber N S 1988 *The Mushrooms Hunter's Field Guide* (USA: Michigan Press)

[5] Simon M 2013 Support of acquisition of recognition of mushroom images *Procs. of Central European Seminar on Computer Graphics (CESCG)* pp 1-8

[6] Sinaga D K 2014 *Implementation of Learning Vector Quantization (LVQ) in Mushrooms Growth Identification* (Bandung: Universitas Komputer Indonesia)

[7] Subramaniam A and Oh B J 2016 Mushroom Recognition Using PCA Algorithm *International Journal of Software Engineering and Its Applications*10(**1**) pp 43-50

[8] Nambi S T and Subha M V 2012 Classification of Stock Index Movement using K-Nearest Neighbor (k-NN) Algorithm J *Wseas Transactions Journal of Information Science and Applications* Vol 9 pp 261-270

[9] Adiwijaya, Winda T and Violina S 2012 *Analysis and Implementation of Zoning and Classification Method of KNN (K-Nearest Neighbor) On Student Grades System with Optical Character Recognition (OCR)* (Bandung: Universitas Telkom)

[10] Chandra P, Jabbar M A and Deekshatulu B L 2013 Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm *International Conference on Computational Intelligence* 10 pp 85-94

[11] Rahmat R F, Chairunnisa T, Gunawan D and Sitompul O S 2016 Skin color segmentation using multi-color space threshold *International Conference on Computer and Information Sciences, ICCOINS 2016 – Proceedings* pp. 391-396