Classification of Zingiber Plants Based on Stomate Microscopic Images Using Probabilistic Neural Network (PNN) Algorithm

U Andayani1*, I B Sumantri2 and B Arisandy1

1 Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
2 Department of Pharmacy, Faculty of Pharmacy, Universitas Sumatera Utara

*Email: ulfi.andayani@usu.ac.id

Abstract Herbal plants are medicinal plants included in the first choice for medicinal ingredients for rural residents. Herbal plants such as ginger and lime are plants that are often used and have plants that are similar in level. One way to identify these types of plants is through stomata microscopic imagery. As for the inspection that is still done manually by the pharmacists, the manual inspection takes a long time and misidentification may occur because there are some similarities with direct viewing, so a method is needed to identify the types of herbal plants based on stomata microscopic images automatically and to improve accuracy in the identification process. The method proposed in this study is Probabilistic Neural Network for the identification of herbal plant types based on stomata microscopic images and the algorithm used to identify these plant types Probabilistic Neural Network (PNN). Before the identification stage, the image will go through three stages, namely preprocessing, segmentation and feature extraction using the Gray level co-occurrence matrix method. After testing using 20 microscopic image data of ginger and bitter ginger stomata. It was concluded that the proposed method has the ability to identify types of herbal plants with an accuracy percentage of 90%.

1. Introduction
In Asian countries, especially Indonesia, China and Korea, herbal medicines are included in the first choice for treatment for rural residents. In developed countries also tend to switch to traditional medicine, especially herbal ingredients. This shows symptoms of a very significant increase.

Classification or taxonomy of plants is a way of grouping based on certain characteristics. In general, biologists use a classification system to classify plants or animals that have similar structures, then each group of plants or animals is paired with a group of plants or other animals that have similarities in other categories. Linnaeus's classification system continues to be used today because of its simple and flexible nature so that a new organism can still be included in the classification system easily. The names used in the Linnaeus classification system were written in Latin because in the days of Linnaeus Latin was the language used for formal education. The classification of living things is based on the similarity in differences in the characteristics and benefits possessed by living things. One example of classification of plants based on their benefits is that plants are grouped into medicinal plants, clothing plants, ornamental plants, food plants and others [1].
Plant morphology not only describes the shape and composition of the plant body, but also has the duty to determine whether the function of each part in plant life, and then also try to find out where the origin of the shape and its composition [5].

About 90% of water released by plants comes out through the stomata. Guard cells function to influence the stomata to open or close. High temperatures and excessive transpiration can also be a cause of closing the stomata during the day [7]. Ginger is an Indonesian spice that is very important in daily life, especially in the health sector. Ginger originates from Asia Pacific which spread from India to China [8].

Rhizome contains essential oils, saponins, flavonoids, and tannins. The main ingredients of essential oils are zerumbone sesquiterpenoids which have biological activities, such as antitumor and antitumor [6].

Several previous studies have been carried out to resolve the identified problems, including those conducted by research [1] Wu, S. G., et all (2007) This research uses Probabilistic Neural Network (PNN) algorithm data that is used 1800 leaves to classify 32 types of plants and obtain an accuracy of 90%.

In further research, [2] using the Probabilistic Neural Network (PNN) algorithm the data used by 20 leaf species from the test results shows that a high error rate occurs when using a spread value of 0.01. and the smallest error rate occurs when using spread 0.25 to 0.5. and accuracy obtained is 89%. In other further research, [3] using the Probabilistic Neural Network (PNN) algorithm the training data is performed by entering refiners in the range [0.001, 0.03] with a step of 0.001. The results of the training carried out using leaf spot data was a smoothing parameter value of 0.025 with an accuracy of the training set of 100% and a validation set of 85.809%.

[4] The study uses the Backpropagation Neural Network (BNN) algorithm and Probabilistic Neural Network (PNN). The data obtained were 600 data for 6 varieties of longan, each of which had 100 data samples. From 100 data samples from each variety, 80 data samples were used as training data and 20 data samples were used as testing data. From the results of the training process, then testing is carried out to see the ability of its introduction. For testing, 1/3 of the data used for each group obtained an accuracy of 91%.

2. The data and method used

2.1. The data used

The image data used in this study is the pharmaceutical biology laboratory of North Sumatra University, the dimensions of the image used are 10x40 with graphic portable network (PNG) format. All data totaling 200 images, where the data sharing is shown in table 3.1

| No. | Microscopic image of Ginger and Bitter ginger stomata | Amount of Data |
|-----|------------------------------------------------------|--------------|
| 1.  | Ginger                                               | 100          |
| 2.  | Bitter ginger                                        | 100          |

The collected images will then be divided into two datasets, the training dataset and the test dataset which will be used to determine the accuracy of the identification process. The division of the entire image into a training dataset and test dataset is done by comparing the training dataset by 90% and the test dataset by 10% of the overall data. Details on the distribution of the dataset can be seen in the table 3.2 and 3.3.
2.2 The method used

The method proposed to identify types of herbal plants based on microscopic images in this study consisted of several steps. These steps start from collecting ginger microscopic image data (*zingiber officinale*) and bitter ginger (*zingiber zerumbet*) which will be used for training data and testing data. Then the preprocessing stage consists of forming a green channel image and image sharpening.

The next stage is the extraction of features from the results of segmentation using the gray level co-occurrence matrix. Then the identification stage is divided into two, namely the microscopic identification of ginger stomata (*zingiber officinale*) and bitter ginger (*zingiber zerumbet*) using Probabilistic Neural Network. After these steps will be obtained from the results of identification of types of herbal plants based on microscopic images of stomata. These steps can be seen in the general architectural form in Figure 1.

| No. | Microscopic image of Ginger and Bitter ginger stomata | Amount of Data |
|-----|------------------------------------------------------|----------------|
| 1.  | Ginger                                              | 90             |
| 2.  | Bitter ginger                                       | 90             |

**Table 3.2. Distribution of training datasets.**

| No. | Microscopic image of Ginger and Bitter ginger stomata | Amount of Data |
|-----|------------------------------------------------------|----------------|
| 1.  | Ginger                                              | 10             |
| 2.  | Bitter ginger                                       | 10             |

**Table 3.3. Distribution of training datasets.**

**Figure 1.** General architecture.
2.2.1 **Input.** The input is an image from a stomata microscopic image dataset, which is an image obtained through the pharmaceutical laboratory of the Universitas Sumatera Utara (USU).

2.2.2 **Pre-processing.** Before the data is used, the data is first processed through the stages of image processing which aims to produce a better image for the next stage of processing. This preprocessing phase consists of forming gray scaling and edge detection on the image (canny edge detection).

2.2.3 **Segmentation.** Stages after preprocessing is segmentation that aims to get the edges of the object by using canny edge detection. Canny edge detection can mark parts of an image detail and improve blurred image details. At this stage the process is carried out utilizing the OpenCV library.

```
Input image to be processed.
Load the Opencv library for the canny process.
Reads the image from the file and saves it to the matrix object.
Read the image.
Create an empty matrix to save the processed image.
Apply the canny function to detect edges.

Imgproc.Canny (source, dest, 100, 200, 3);
Imgproc.Canny (an object representing the source (input image) for this operation, a Mat object representing the destination (edge) for this operation, a double type variable representing the first threshold for a hysteration procedure, a double type variable representing a second threshold for a hystetration procedure) ;
Create and save image processing results.
```

**Figure 2.** Pseudocode for the canny edge detection process.

![Image of canny edge ginger detection results.](image)

**Figure 3.** Image of canny edge ginger detection results.

2.2.4 **Extraction features.** After the image processing is carried out the next stage is extracting features or features from the final stages of image processing. Feature extraction in this study uses the initial step method to get the moment invariant value by calculating the moment value of the image. The moment value sought is the final result of the image processing stage that is canny edge detection measuring 100x400 pixels. Calculation of the moment value is done using equation 1. The moment value sought is $m_{00}, m_{10},$ dan $m_{01}$ for every object that exists.

$$m_{pq} = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} x^p y^q f(x, y)$$ (1)
The moment value obtained using equation 1 is as follows.

\[
\begin{array}{ccc}
m_{00} & m_{10} & m_{01} \\
695217.0 & 2.8890943E7 & 1.58404658E8
\end{array}
\]

After the moment value \(m_{00}, m_{10}, \) dan \(m_{01}\) obtained, then the next step is to calculate the central moment value using equation 2.

\[
\mu_{pq} = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} (x - \bar{x})^{p} (y - \bar{y})^{q} f(x,y)
\]

where:

\[
\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}
\]

Central moment value obtained from equation 2, where is the value of \(\bar{x} = 41.55672689246667\) and \(\bar{y} = 227.8492298088223\) are as follows.

1. \(\mu_{11} = 2.735461699994162E8\)
2. \(\mu_{20} = 4.346951030831737E8\)
3. \(\mu_{02} = 9.719116756570135E9\)
4. \(\mu_{30} = 7.276289957362335E9\)
5. \(\mu_{03} = -9.942218400349191E11\)
6. \(\mu_{12} = -1.137597457590645E10\)
7. \(\mu_{21} = -2.348558667746482E9\)

After the central moment value \(\mu_{11}, \mu_{20}, \mu_{02}, \mu_{30}, \mu_{03}, \mu_{12}\) and \(\mu_{21}\) obtained next will be carried out the process of ginger center value using the moment equation 3.

\[
\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}
\]

where:

\[
\gamma = \frac{p+q}{2} + \frac{1}{\mu_{00}} = m_{00}
\]

From the results of the ginger moment the center \(\mu_{11}, \mu_{20}, \mu_{02}, \mu_{30}, \mu_{03}, \mu_{12}\), and \(\mu_{21}\) in equation 3 the following values are obtained.

1. \(\eta_{11} = 5.659653873172104E-4\)
2. \(\eta_{20} = 8.993815646619662E-4\)
3. \(\eta_{02} = 0.02010879435645232\)
4. \(\eta_{30} = 1.80554800019766295E-5\)
5. \(\eta_{03} = -9.782271731631517E-4\)
6. \(\eta_{12} = -2.8228490462068607E-5\)
7. \(\eta_{21} = -5.827743856996758E-6\)

2.2.5 Identification. After the feature extraction value is obtained, the next stage is identification using the Probabilistic Neural Network (PNN) method. The stages in this method are divided into two stages, namely the training and testing phase. At the training stage the value extraction value used is the value of the training data. At the testing stage the feature extraction values are obtained from the test data which will be entered into the pattern layer then to the summation layer and the highest probability values will be grouped into that class.

The training process in PNN is carried out by steps, where the weight value of each neuron in the pattern layer will be formed by the characteristic vector of the feature extraction in each training data. PNN training process.
Figure 4. Pseudocode the PNN method training process.

1. Whereas the testing process is carried out with the following steps:
2. Enter the test data first.
3. Then the test data will be calculated proximity to the weight vector in the database. Then the kernel gaussian function.

\[
F_{ij}(x) = \frac{1}{2\pi^{d/2}\sigma^d} \exp \left[ -\frac{\|x-x_{ij}\|^2}{2\sigma^2} \right] 
\]  

(4)

The process is done by checking the training data with the category of microscopic image of ginger stomata (*zingiber officinale*) with the amount of data 90, up to microscopic image of stomata bitter ginger (*zingiber zerumbet*) 90 data.

The value of \(x\) is the distance from the test data with the weight vector, while \(y\) is the result of the pattern layer process and the result PNN is the result of the kernel gaussian function 4. Furthermore, after the proximity of the test data distance and weight vector are obtained then the results of the kernel gaussian function are added to the same class, then averaged by the amount of test data corresponding to each class. This process aims to find the probabilities of each class that are carried out using equation 5.

\[
g_i(x) = \frac{1}{2\pi^{d/2}\sigma^d N_i} \sum_{j=1}^{N_i} \exp \left[ -\frac{\|x-x_{ij}\|^2}{2\sigma^2} \right] 
\]  

(5)

The results of calculations on equation 5 can be seen in Fig 3.

Val : 0.10225616740385433
Val1 (value ginger class) : 3.1699411895194842
Val2 (value bitter ginger class): 9.20305506634689
Probability value of ginger and bitter ginger: [7.348324219615403E-4, 0.1135680568512672]

Figure 5. Calculation results on summation layer.

The value of \(val = 0.10225616740385433\) is the result of the addition of functions of the kernel with the same class. Value of \(val1 = 3.1699411895194842\) obtained from multiplying the value of \(val\) with the amount of training ginger data. Value of \(val2 = 9.20305506634689\) obtained from the multiplication of \(val\) values with the amount of training data in bitter ginger. The results of these calculations are then averaged to obtain the probabilities of each class, namely the ginger and bitter ginger probability values [7.348324219615403E-4 PNN values for the ginger category] and [0.1135680568512672 PNN values for the fracture category]. The highest probability value will then be entered into the class which is done by applying the Bayes’s decision function. The results of the tests carried out issued slime output. PNN architecture from the identification process that has been done can be seen in figure 6.
2.2.6 Output. Output in the form of data display on the system that produces the final classification whether the initial input image is ginger or bitter ginger.

3. Conclusion

3.1. Conclusion

Based on the application of methods from general architecture designed to identify types of herbal plants based on stomata microscopic imagery, the following conclusions are obtained:

The proposed method is able to identify types of herbal plants based on stomata microscopic imagery using 90 training data and 10 testing data which results in an accuracy of 90%.

Reference

[1] Soepomo.1987. *Morfologi Tumbuhan*. Yogyakarta: PT.Gajah Muda University Press.
[2] Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL.2007. *A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network*.arXiv [CS.AI] 1 : 4289.
[3] Gregorius Satia Budhi ,Tok Fenny Handayani, Rudy Adipranata. 2008. *Aplikasi Pengenalan Daun Untuk Klasifikasi Tanaman Dengan Metode Probabilistic Neural Network*. Auditorium Universitas Gunadarma.
[4] H. Syahputra and A. Harjoko, 2011, *Klasifikasi Varietas Tanaman Kelengkeng Berdasarkan Morfologi Daun Menggunakan Backpropagation Neural Network dan Probabilistic Neural Network,”* *IJCCS*, vol. 55281, pp. 11–16.
[5] Tjitrosoepomo. 2005. *Morfologi Tumbuhan*. Yogyakarta: PT.Gajah Muda University Press.
[6] Kadir, Abdul. et al. 2013. *Teori dan Aplikasi Pengolahan Citra*, Yogyakarta: Andi.
[7] Mulyani, Sri. 2006. *Anatomi Tumbuhan*. Yogyakarta: Kanisius.
[8] Paimin, F.B., dan Murhananto. 2002. *Budidaya, Pengolahan, dan Perdagangan Ginger*. Jakarta: PT. Penebar Swadaya.
[9] Setiawan, W., Utoyo, M.I. and Rulaningtyas, R., 2019. Classification of neovascularization using convolutional neural network model. *TEKOMNIKA, 17*(1), pp.463-472.