International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Develop Management Information System of Medicinal Plants Based on Plant Image Identification Techniques using Neural Network

Le Trieu Tuan

University of Information and Communication Technology – Thai Nguyen University,
Quyet Thang District, Thai Nguyen Province, Vietnam

Abstract: ***This paper presents the use of cellular neural networks in image processing to develop a management information system data medicinal plants of Thai Nguyen province. With this technique, the input leaves medicinal plants will be identified via propagation techniques, from which the system will detect image input herbal leaf tree species. Users will implement lookup information on medicinal plants that you care about through a complete identification system. The system will bring convenience to users, especially when people want to pay attention to information about medicinal plants. With the development of this system will contribute to the management, the pharmaceuticals can be a useful tool for monitoring, statistics, identification and management of medicinal plants effectively.***

Keywords: Cell neural network; Medicinal plants data; Back propagation; Management Information System; Leaf Recognization.

1. Introduction

Medicinal plants are one of the most important plants in the life of everyone, especially meaningful in health care. Medicinal plants have many different uses and characteristics: plants are effective for treating people, animals, plants used for wood, plants for food, etc. And based on the characteristics, utility today, one can classify into hundreds of thousands of different medicinal herbs [1].
With the use and high economic value of medicinal plants, the exploitation of more and more, while the planted area or the natural development of medicinal plants is not enough to offset the exploited. In addition, the illegal exploitation, theft has led to loss of many sources of medicinal plants precious. As a result, medicinal plants are being harvested more and more, leading to exhaustion [2]. In the world, there are many works by several authors who carry out research on plant material, especially the field of identification based on plant characteristics and serve to improve the efficiency of tree harvesting and management medicine. We can include: Group of authors Wu, Zhou and Chaonan Wang [3]; Valliammal and Geethalakshmi [4]; Bhardwaj, Kaurand and Kumar [5] collects and manages medicinal plants by building a leaf database based on the characteristics of the leaf morphology and morphology (such as mass, indentation, leaf margin, leaf vein, etc.) [6]. Ehsanirad and Kumar [7] manages tree databases using image processing technology to classify trees based on leaves. Extract the leaf characteristics, using the gray level matrix. Du, Wang and Zhang [8], Lee and Hong [9], constructed a leaf database based on the geometry and leaf boundaries for subsequent classification and lookup. Chaki and Parekh [10] built a leaf-based tree identification system and built a database of trees with 180 species and divided into 60 species in a single layer to manage tree data and serve for the investigation assist; Wu and his colleagues built a database of 1800 species and classified it into 32 classes.

The authors of Prasad Babu and B. Sinivasa Rao have developed a model for identifying the disease from the risk of harm or disease management. Using this method, users can scan an image of a diseased leaf to identify which leaf is the species, the risk of disease or the way it is cured and possibly get a solution for the problem. Care of it to protect, maintain and develop medicinal plants. So the world has studied the proposed solutions for the exploitation and management, as well as the database of medicinal plants and classifies them based on the characteristics In terms of shape, color, especially their use for humans. Based on this fact, this paper presents a technique for the development of management information systems using neural networks to support the monitoring, management and protection of precious medicinal plants.

2. Method and data for the system

Based on the research on the problems of medicinal plants such as collecting, managing, monitoring, identifying and protecting them from works published in the country and in the world. . Especially, through the survey of practical needs and experiments, we have introduced the process of collecting photos of medicinal leaf and photo pre-treatment as follows:

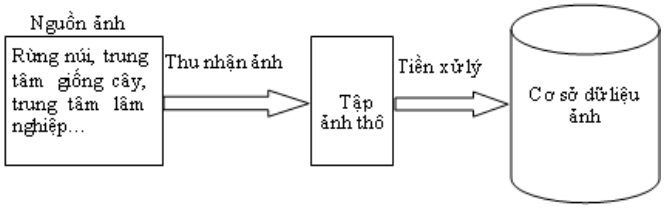

Figure 1: Diagram of the system for collecting and treating medicinal plants

Photos of medicinal plants are collected through specialized photographic receivers such as cameras, digital cameras, scanners, etc. But the picture obtained at this step has a lot of

noise that makes the receiving process. Image manipulation and complex management slow down the system, so it needs to be processed before identification, this step is called pretreatment.

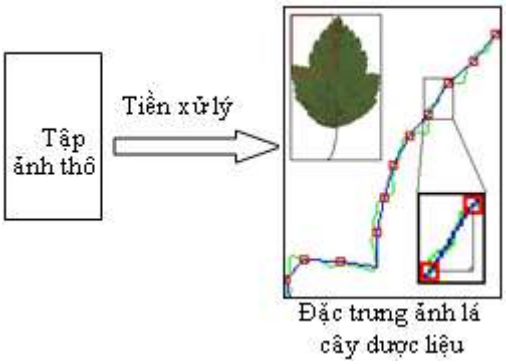

Figure 2: Image data pre-processing

3. Identification of medicinal plants using neural cell networks

The reverse propagation algorithm is divided into two distinct stages: in the first step, the derivative function must be calculated according to the weights. At this point the error will be propagated back to the network. Second, the derivatives will be used to calculate the adjustment for the weights. And the simplest technique used here is the gradient descent technique.

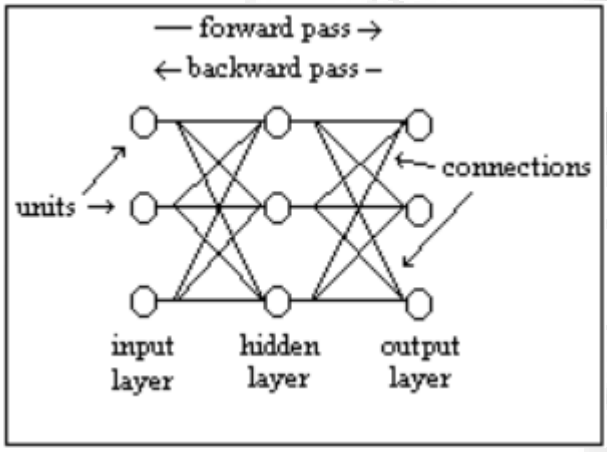

Figure 3: Backward propagation using gradient descent

In spreading networks, each node generally aggregates its inputs according to the following formula:

$$a_j = \sum_i w_{ji} z_i \quad (1)$$

For z_i is the input value, or the output value of a node connected to the node j and w_{ji} is the weight of the connection. This sum is calculated on all nodes connected directly to node j. We know that the threshold of a node is also included in the sum by adding a fixed value of 1. The sum in (1) is transformed by a nonlinear function output the output value j of node j according to the formula:

$$z_i = g(a_j) \quad (2)$$

We now need to determine the value of the weights in the network by minimizing the error function. Here we will consider the error function as a sum of all errors in each sample. This amount will be calculated on all samples of the training set:

$$E = \sum_n E^n \quad (3)$$

Where n is the label of each sample.

We also assume that the error E^n can be expressed as a function of the output variables, that is:

$$E^n = E^n(y_c, \ldots, y_c) \quad (4)$$

Our goal here is to find a function that calculates the derivative of the error function according to the weights and thresholds of the network. For each sample, we will assume that we have provided an input vector corresponding to the input and calculated the output values of the hidden nodes as well as the output nodes of the formulas (1), (2). This process is often referred to as the spreading process in the network.

Now consider the derivative of E^n with the wji weights. The output value of the nodes will depend on each input n type. However, to make it easy to see, we will ignore the n character encoding on the input and output variables. We first need to note that E^n depends on the wji weighting through the total input ai of node j. So we can give the formula for the partial derivatives as follows:

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j} * \frac{\partial a_j}{\partial w_{ji}} \quad (5)$$

From (1) we have:

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \quad (6)$$

Thus it follows:

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j z_i \quad (7)$$

Where:

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} \quad (8)$$

From formula (8) we find that for calculating the derivative we only need to compute the value for each hidden node and node in the network. With the output node, the calculation of δ_k is very simple. We have:

$$\delta_k \equiv \frac{\partial E^n}{\partial a_k} = g'(a_k)\frac{\partial E^n}{\partial y_k} \quad (9)$$

To compute (9) we need to find the formula g '(a) and $\frac{\partial E^n}{\partial y}$.

To calculate δ for hidden nodes, we need to use the formula of partial derivatives:

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k}\frac{\partial a_k}{\partial a_j} \quad (10)$$

Where the sum is calculated on the nodes to which node j is connected. The arrangement of the nodes as well as the weights is illustrated in Figure 4.

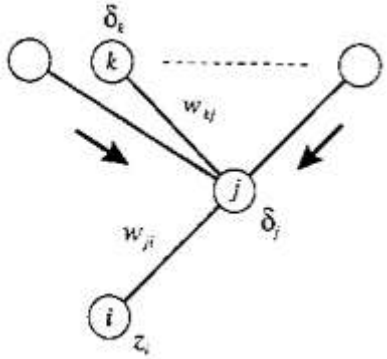

Figure 4: Illustrates the calculation of δ_j for the computation of the hidden node j

$$\delta_j \equiv g'(a_j)\sum_k w_{kj}\delta_k \qquad (11)$$

This formula states that the value of δ for a hidden node can be calculated from the backward propagation of the δ values of the higher hidden nodes in the network, as illustrated in Figure 5.

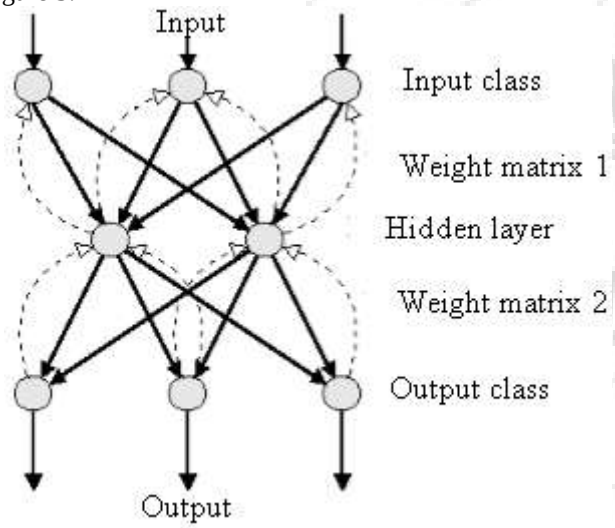

Figure 5: Neural model in the sample recognition problem

The neural network in the system will consist of three layers: the input layer, the hidden layer and the output layer. The hidden layer can also be divided into several smaller hidden layers but in this system we will only use a hidden layer to speed up the calculation when installing the system. The input of the neural network is the characteristic of the leaf image. A feature of the leaf image consists of two parts, cos and sine of an angle in the above triangle. The number of input layers is calculated by the number of points divided by 2. The output of the neural network will be equal to the number of species that the system can discriminate.

4. Results and Discussion

We'll use a set of 70 images of 10 species of leaves to test the accuracy of the system. The number of photos in the training set and test series of each species that is shown in Table 1:

Table 1: Number of photos in the training set and pilot set of each species

Species	*Number of training photos*	*Number of test images*	*Total number of photos*
Averrhoa carambola	7	3	10
Bryophyllum	7	3	10
Melastoma	7	3	10
Mentha	7	3	10
Citrus grandis	7	3	10
Morus alba	7	3	10
Houttuynia	7	3	10
Bougainvillea	7	3	10
Piper lolot	7	3	10

5. Conclusion

Successful establishment of the leaflet information management system successfully solves the problem requirements such as the ability to find image characteristics and identify the trained leaf samples. The results of the system are relatively accurate. In the future, in order for the system to achieve optimal accuracy, additional attributes related to the leaves and increasing the amount of data and input quality of the system should be added. Although the system is simple but has solved the problem of leaf identification with relative accuracy. In order to improve the system, it is necessary to increase the size of the identification training data set and the data to be diversified. In addition, the system will also be developed and expanded to provide more useful information on leaves and plants for users such as: What is the leaves of the family, living mainly in which regions, or systems, can grow in the direction of identifying and classifying diseased leaves based on the identity of the diseased leaves. These applications are really useful and serve the daily tasks in human life.

References

[1] Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation, 2007, Vol. 185, No 2007, pp. 883-893.

[2] Jyotismita Chaki and Ranjan Parekh, "Plant Leaf Recognition using Shape based Features and Neural Network classifiers", International Journal of Advanced Computer Science and Applications, 2011, Vol. 2, No.10.

[3] Kue-Bum Lee, Kwang-Seok Hong, "Advanced Leaf Recognition based on Leaf Contour and Centroid for Plant Classification", *International Journal of Bio-Science and Bio-Technology*, 2012, Vol. 5, No.2.

[4] N.Valliammal and Dr.S.N.Geethalakshmi, "Automatic Recognition System Using Preferential Image Segmentation For Leaf And Flower Images", *Computer Science & Engineering: An International Journal (CSEIJ)*, 2011, Vol.1, No.4

[5] M. S. Prasad Babu & B.Srinivasa Rao, *Leaves Recognition Using Back Propagation Neural Network-Advice for Pest & Disease Control On Crops*, Technical Report, Department of Computer Science and Systems Engineering, Andhra University, India, 2007.

[6] Qingfeng Wu, Changle Zhou and Chaonan Wang, *Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network*, Published by the Center for Computing Research of IPN, Mexico,2006.
[7] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, *A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network*, Published at IEEE 7th International Symposium on Signal Processing and Information Technology, Cario, Egypt, 2007.
[8] D. Heckenberg, B. C. Lovell, "A Gesture Driven Computer Interface", *Proceeding of Visual Communications and Image Processing*, 2000, SPIE, 4067, 261-268.
[9] J. Mathews, "An Introduction to Edge Detection: The Sobel Edge Detector", 2002.
[10] N. Sakai, S. Yonekawa and A. Matsuzaki, *Two-dimensional image analysis of the shape of rice and its applications to separating varieties*", Journal of Food Engineering, vol 27, 1996, pp. 397-407.

Author Profile

Le Trieu Tuan received the B.S. in Information Technology from Thai Nguyen University in 2010 and M.S. degrees in Information Technology from Enverga University in 2013. He now with lecture, University of information and communication technology – Thai Nguyen University.