LEAF RECOGNITION SYSTEM UTILIZING ARTIFICIAL NEURAL NETWORK FOR MEDICINAL PROPERTIES FAMILIARIZATION OF THE INDIGENOUS AND CULTIVATED PLANTS AND TREES IN THE PHILIPPINES

MARCELO R. ANZANO[1], ANGELICA MAE A. BANAAG[1], CLAIRE T. CAPIO[1], LARA MAE D. EDLES[1] AND ORLAND D. TUBOLA[1,2]

[1]Department of Computer Engineering, College of Engineering, Polytechnic University of the Philippines
[2] Center for Engineering and Technology Research, Institute of Science and Technology Research, Polytechnic University of the Philippines

Abstract - Plants and trees are important not only for their oxygen production ability but also for their medicinal properties. This study about leaf recognition plays a very important role in the field of education especially for those who are interested in the medicinal uses because it can provide an easier way of learning the significance of different species of plants and trees in the Philippines. The researchers of this study created artificial neural network system that will recognize a leaf and will give the user information about it. Artificial neural network is an information processing system that is based on how brain learns and process information. Like the brain, neural network learns by feeding information to its database and training its neurons to function properly. The neural network will then recognize the input by matching it to its trained database. The researchers used two different data base, one containing 300 samples and the other with 100. The system was tested with the use of different numbers of neurons and was found to be most accurate with 75 neurons in the database containing 100 samples as it shows an accuracy of 99%.

Keywords: *artificial neural network, leaf recognition, information processing system, plants and trees information.*

1. INTRODUCTION

In this study, leaf recognition was designed for the user to be familiar with the plants and trees in their community and to be knowledgeable about their medicinal uses. In particular, its educational values would immensely contribute to elementary and high school students who are interested with this kind of information. According to (Ahmed, 2011) medicinal plant is any plant which, in one or more of its organs, contains substances that can be used for therapeutic purposes. It is designated as medicinal if it is implied as a useful drug or therapeutic agent or an active ingredient of a medicinal preparation. Since time immemorial, medicinal plants were highly utilized by many people in every part of the world. It is particularly popular in the Philippines especially in rural areas.

This study's database includes the scientific information and medicinal uses of plants and trees' leaves in the Philippines. In order to recognize the leaves properly and

accurately the researchers developed an image recognition algorithm that matches the leaf's photograph against the leaf-image database using the points taken in the numerous parts of the leaf.

The best matches were then ranked and returned to the user for final verification. Leaves from different species are sometimes quite similar that is why one of the main technical challenges in using leaves to identify plant species is to find the effective representations of their shapes that capture their most important characteristics. In order to make an effective algorithm for this study the researchers made use of artificial neural networks.

The Artificial Neural Network is a computational model which is capable of pattern recognition as well as machine learning. Machine learning is the ability of the machine to generate an output based on what is learned from the data rather than following a certain programmed algorithm (Wernick *et al*., 2010). Pattern recognition is almost the same with machine learning; it's one of the many branches of artificial intelligence that focuses on the regularities in the data (Bishop, 2006). The artificial neural network is based on the biological neural network. This means that like the biological neural network, this artificial neural network also have neurons that activate when a signal is received. The neurons in the artificial neural network are trained with the use of input and target data (Gershenson, 2003).

There are several categories on how artificial neural network works. The categories are input-output curve fitting, pattern recognition and classification, clustering, and dynamic time series. This study focuses more on pattern recognition and classification for the system to recognize the leaf based on a given image. Classification is the process of identifying which set of data belongs to a certain category (Alpaydin, 2010). The data is classified through recognizing the regularities found in the data.

This system is created through the use of what is stated on the previous paragraphs. An artificial neural network will be utilized in this system. The type of the neural network created is feed – forward. Meaning, the information moves in one direction only, forward, from the input through the neurons, then to the output. Neurons in this system are trained based on what is encoded in the database. It falls on the category of classification and recognition for the reason that the leaf is recognized by classifying the data collected from the input image. And when the leaf is recognized based on what class it belongs, the details will be fetched on another database by matching the class number and the columns.

2. METHODOLOGY

2.1 Gathering of images and information of leaves

The images of the leaves in the database are scanned pictures of actual leaves gathered by the researchers from actual trees around Metro Manila and Antipolo. The information included in the database was taken from references listed in the latter part of this paper.

2.2 Creating the database

The system is created and implemented with the use of Matlab. Before the algorithm of the program is created, it is very important to setup first the database. To do this, three Matlab variables were created; these are the input, target, and the details. These variables are matrices with different dimensions and since the accuracy of the system is important, there are two sets of these matrices created in case a problem with accuracy occurs. The issues regarding the accuracy of the system is due to the limited number of sample inputs and the leaves' shape condition. The input's dimensions are 10x328 and 10x105, the target's dimensions are 1x328 and 1x105, and the details' dimension is 12x66.

Table 1. Part of the input variable containing ten points from three different leaf images.

LEAF	1	2	3
1	0.9944	0.9981	0.9979
2	0.7085	0.7664	0.65759
3	0.1029	0.0050	0.0046
4	0.2376	0.0059	0.1183
5	0.3292	0.0062	0.1800
6	0.3143	0.0072	0.1687
7	3.0541	1.4339	36.8049
8	2.4170	2.4609	2.3670
9	0.3689	2.2624	2.7760
10	0.0010	0.0033	0.0037

Table 2. Part of output variable showing the class number where the leaf image belongs to.

LEAF	1	2	3
1	1	2	2

As for the details of the variable, it is a matrix with 12x66 dimensions. One column is dedicated to one kind of leaf, and each row of it is dedicated to different details like its common name, scientific name, etc. It is very important that the class number set on a specific type matches the column number where the details are encoded.

Table 3. Part of details variable showing information about the leaves

	1	2	3
1	'Common Name: Alagaw '''	'Common Name: Alugbati '	'Common Name: Ampalaya'
2	'Scientific Name: Premna odorata '	'Scientific Name: Basella alba '	'Scientific Name: Momordic...
3	'Kingdom: Plantae'	'Kingdom: Plantae '	'Kingdom: Plantae'
4	'Division: Magnoliophyta'	'Division: Tracheophyta'	'Division: Tracheophyta'
5	'Class: Magnoliopsida'	'Class: Magnoliopsida'	'Class: Magnoliopsida'
6	'Order: Lamiales'	'Order: Caryophyllales'	'Order: Cucurbitales'
7	'Family: Verbenaceae'	'Family: Basellaceae'	'Family: Cucurbitaceae'
8	'Genus: Premna L.'	'Genus: Basella L.'	'Genus: Momordica L.'
9	'Species: Premna odorata Blanco'	'Species: Ceylon spinach'	'Species: balsampear'
10	'Locality: Mt. Banahaw'	'Locality: Found in settled an...	'Locality: It is mostly cultivat...
11	'Occurring in two forms in the P...	'Locality: Throughout the Phi...	'''Locality: Throughout the P...
12	'''Medicinal Uses: (a) Leaf Leaf de...	'''Medicinal Uses: (a) Pulped l...	'''Medicinal Uses: (a) Decoct...

2.3 Collection of points from the leaf images

Ten points from the image is collected, which will be the ones to be encoded in the input variable. To obtain the ten points from the system, an algorithm was created which will also be used when the algorithm of the whole system is created. The ten points that this algorithm will get is located around the shape of the leaf, it will then be stored in a temporary variable which is an array with 1x10 dimensions. When setting up the input variable is done, the next thing to do is to setup the target variable. The target variable must contain the class numbers which a leaf image belongs to.

2.4 Artificial neural network using Matlab

The artificial neural network was created through the use of the neural network tool in Matlab. First, the two variables, input and target, are loaded to the workspace. Another variable which will be named as sample is created which is a 10x1 array and also loaded to the workspace. Its values are the same with the first column of the input variable. This variable isn't really a part of the system. It is created for testing purposes only. The neural network tool is opened and the three variables are imported there from the workspace afterwards.

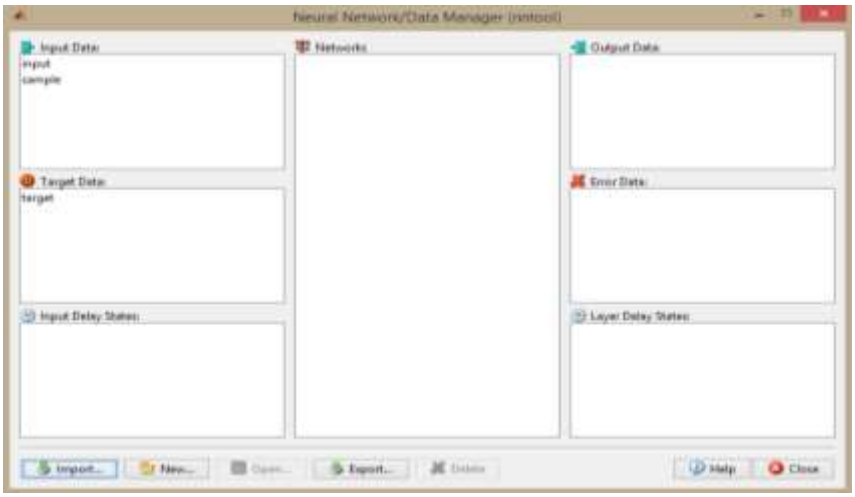

Figure 1. Input, sample, and target imported to the neural network tool of Matlab.

2.5 Training of the created artificial neural network with the use of the database.

Several artificial neural networks were created with 10, 50, 75, 150, 250, and 300 neurons and use the most common type of network which is the feed – forward back propagation. This network will be trained through the use of the input and target variable. After several trainings, the network is simulated to see how accurate the result is through the use of the sample variable. And if the result is not so accurate, the network is trained again. When the network yields an accurate result already, then it should be saved in the same folder where the other components for this system were saved.

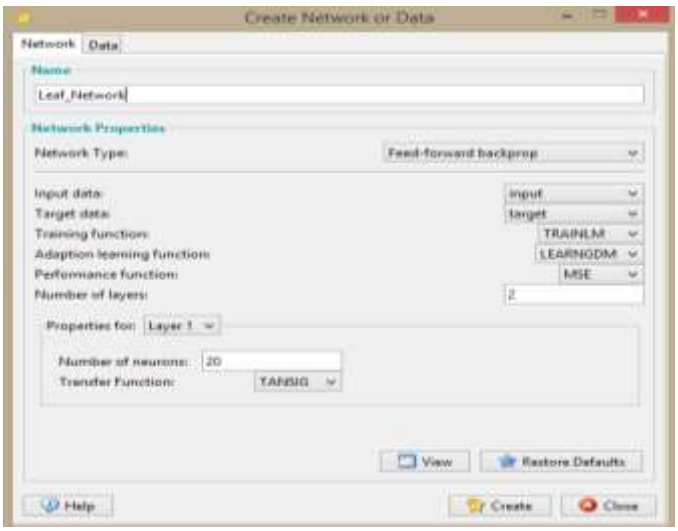

Figure 2. Creating the artificial neural network with the described specification.

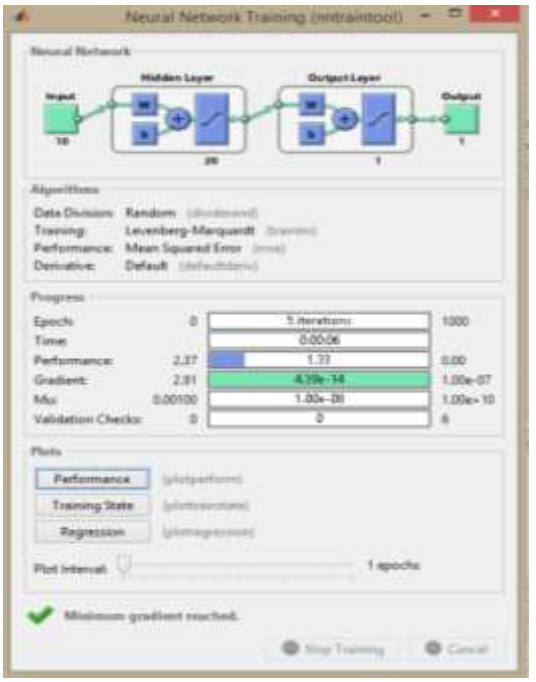

Figure 3. Training of the network.

2.6 Loading of data to the artificial neural network

When the database is complete and the artificial neural network already yields accurate result, it's time to proceed in creating the main algorithm for the system. The algorithm in getting the leaf data is also used in this main program. It is located in the first part of the program followed by the transposition of the variables from a horizontal orientation to vertical and the artificial neural network. The values collected from the image will go through the process of detail fetching. This is done by using the principal component analysis. Then, it will enter the artificial neural network system in which the recognition of the input leaves' image begins.

2.7 Classifying of input data and matching of details

Upon loading of data to the artificial neural network, the network will identify the class number where the values belong to or the most possible class where the data may belong. This class number will serve as the reference when getting the leaf details. Each class number will be representing one column in the details variable, which is why when the class number output is 1, the details located in column 1 is fetched.

2.8 Displaying of image and the different information of the identified leaf using a GUI

Finally, after the recognition and fetching of data, the GUI or graphical user interface is created using MatLab. The GUI consists of two buttons for selecting image and starting the recognition system. The selected image and the recognition result will also be displayed in the GUI. In the recognition result, it will display the common name, scientific name, kingdom, subkingdom, division, infradivision, class, order, family, genus, species, medicinal uses and locality of the recognized plant.

3. RESULTS AND DISCUSSION

This study was conducted for the sole purpose of recognizing leaves with the use of neural networks. Once the image of a leaf has been selected and loaded to the system, the neurons will classify it according to the class number assigned to the leaves indicated in the database. Upon classifying, the system will fetch the data from the details variable then it will display the information of the recognized leaf.

The neural network system created with the use of the database containing more than 300 leaves recognizes the leaves inaccurately no matter what number of neurons were used compared to the neural network system created with the use of the database containing more than 100 leaves trained through the use of 75 neurons. The system was able to recognize the leaf more accurately with its scientific information, medicinal uses and locality. It was also able to accept different sizes of captured frames because there is a part of the algorithm that can resize the image. For further visualization of the result, refer to the figures shown below.

Figure 4 shows the training of the neural network with 75 neurons trained with the database with more than 100 leaves. It shows the details of the training which will be discussed further in this study.

Figure 5 shows the training performance plot from the neural network's training results showed in figure 4. The circle indicates where the neurons' performance are at their best. The lower it is located in the graph, the higher the accuracy of the prediction will be for this plot shows the errors of the neurons.

Figure 6 shows the regression plot of the neural network used for the system. As seen in the plot, there is a broken line, a solid line with a different color, and circles. The circles represent the data, the solid line represents the fit, and the broken lines represent the target. When the fit line as well as the data are aligned with the target, the neural network will be able to yield an accurate result. However, when the data is still scattered and the fit line is not aligned with the target, it means that the neural network will not yield an accurate result.

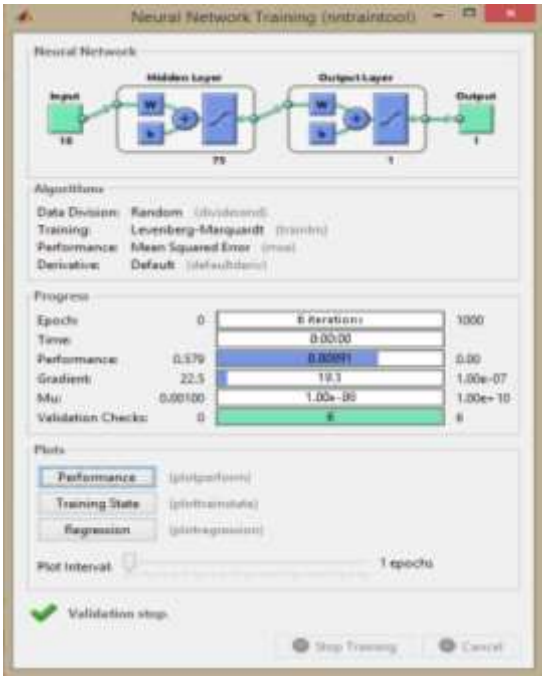

Figure 4. Neural network training with 75 neurons trained with the database containing more than 100 leaves.

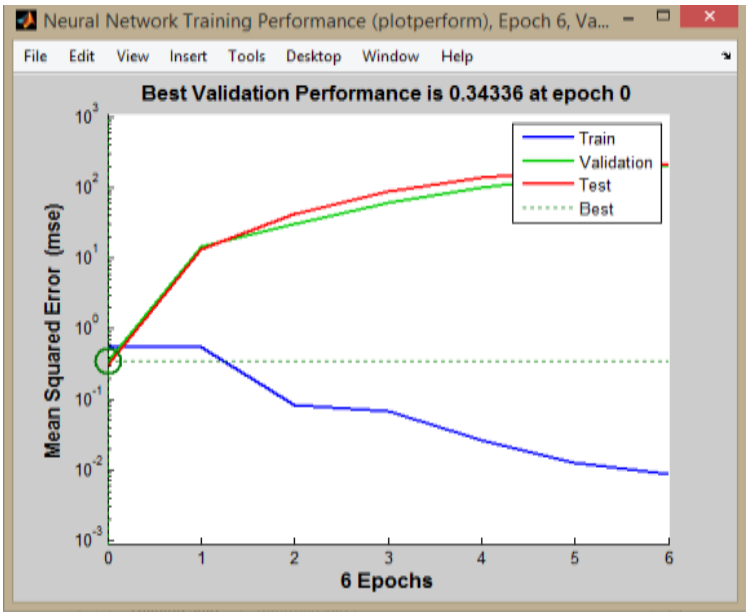

Figure 5. Neural network performance plot with 75 neurons trained with the database containing more than 100 leaves.

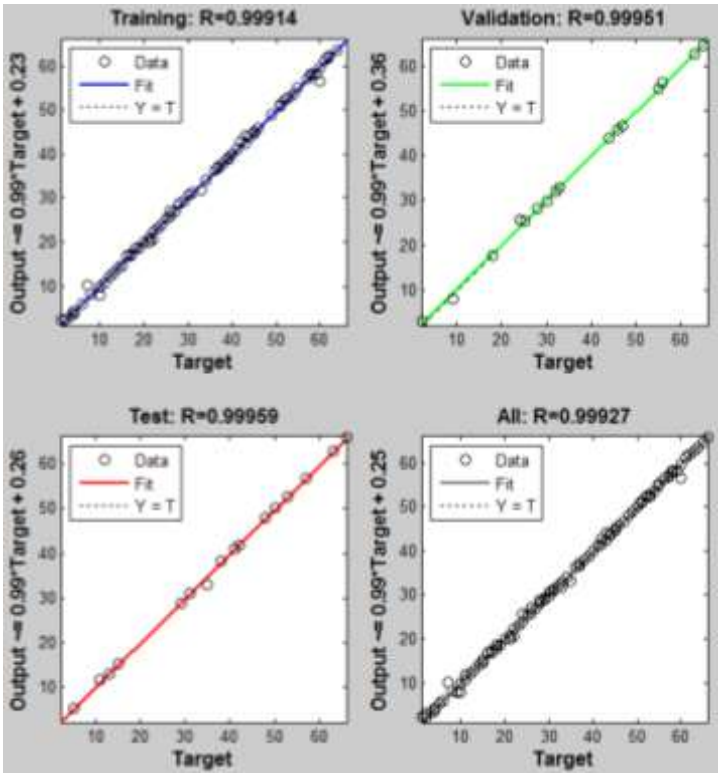

Figure 6. Neural network regression plot with 75 neurons trained with the database containing more than 100 leaves.

Figures 7 to 12 show recognition results from the neural network with 75 neurons trained with the database containing more than 100 leaves. This combination of number of neurons and size of the database is proven to be the most efficient among the other trials conducted in this study. It is tested with Suha leaf, Lubi-lubi leaf, Gabi leaf, Guyabano leaf, Kalamansi leaf, and Langka leaf.

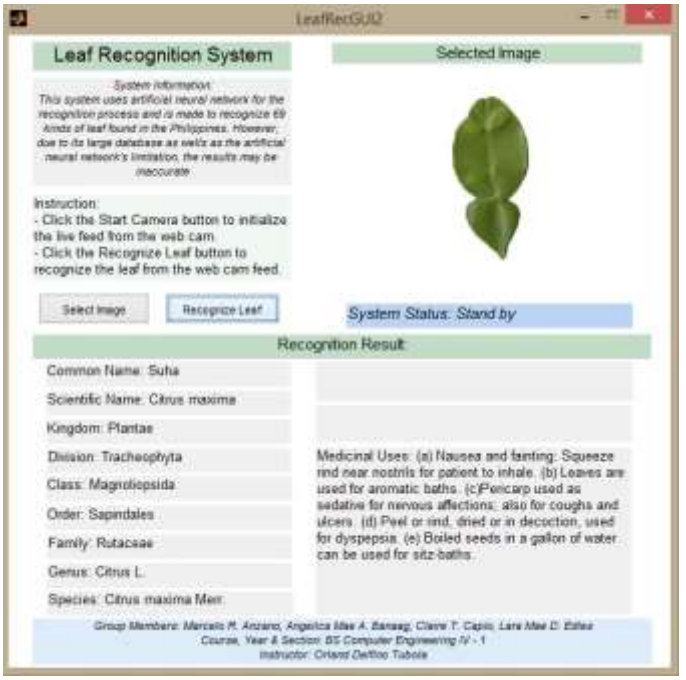

Figure 7. Recognition of Suha leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves.

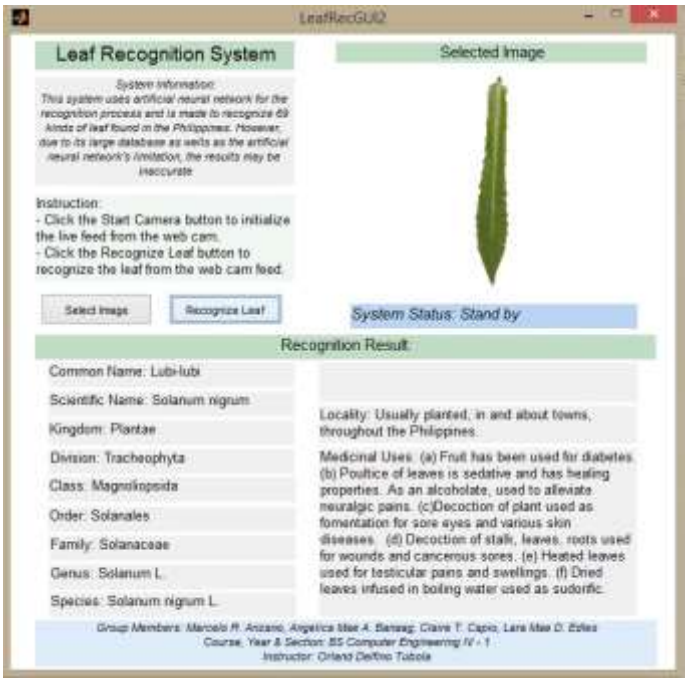

Figure 8. Recognition of Lubi-lubi leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves.

Figure 9. Recognition of Gabi leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves.

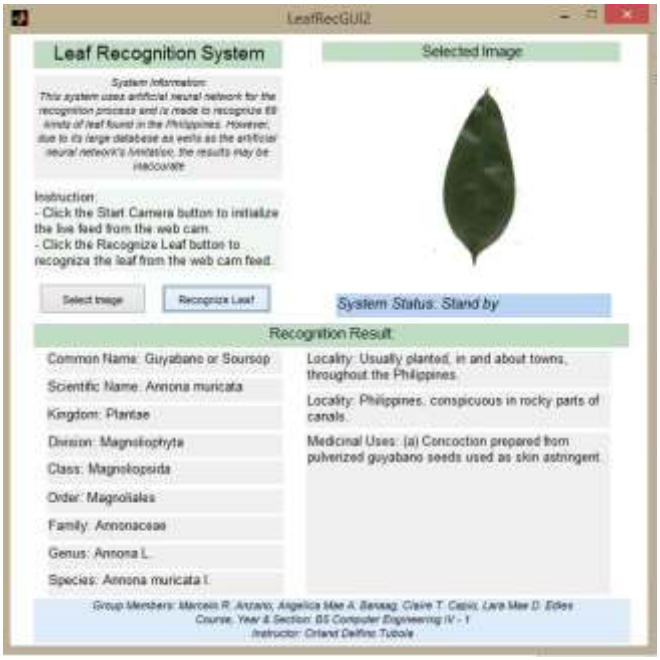

Figure 10. Recognition of Guyabano leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves.

Figure 11. Recognition of Kalamansi leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves

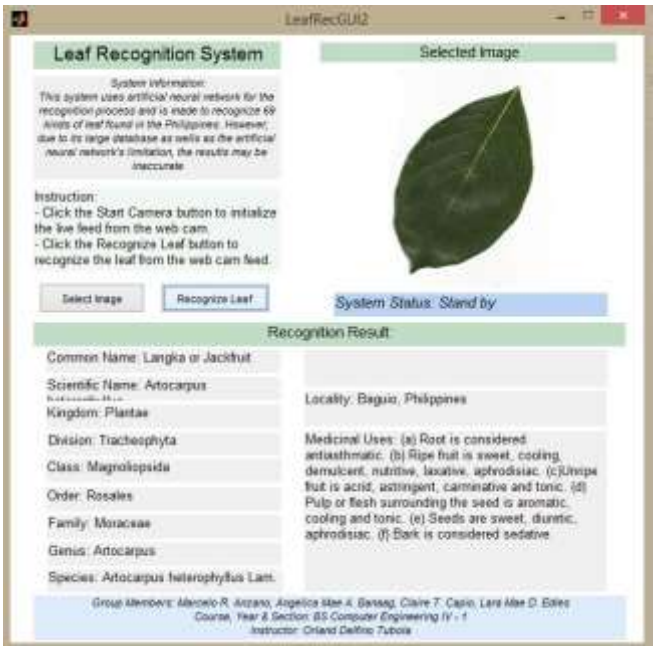

Figure 12. Recognition of Langka leaf using the neural network with 75 neurons trained with the database containing more than 100 leaves.

As stated earlier, several neural networks with different combinations of database and number neurons were created, trained and tested. The table below shows the accuracy of each network. It shows that the neural network with the 75 neurons trained with a database containing more than 100 leaves gives the highest accuracy followed by the neural network with the 50 neurons trained with a database containing more than 300 leaves. For this reason, the neural network with the 94% accuracy can recognize some of the leaves accurately but it cannot recognize accurately as many leaves as the one with the 99% accuracy.

Table 4. Accuracy of each neural network trial.

Number of Neurons	Actual	Prediction	Accuracy
More than 300 leaves in database			
10	20	14.4519	62%
50	20	18.8954	94%
150	20	14.8905	66%
250	20	10.5318	10%
300	20	12.2114	36%
More than 100 leaves in database			
10	20	22.1516	90%
50	20	21.9505	91%
75	20	20.1028	99%

The numbers encoded in the actual column is the actual identification number of a leaf. In this case, the number 20 is the identification number of the Guyabano leaf. The networks were simulated and the results were gathered which can be found in the prediction column. The percentage of accuracy was determined through the use of the percentage error formula.

4. CONCLUSIONS AND RECOMMENDATIONS

In this study, the simulation shows that upon capturing frames of the leaf for recognition, the system displays the correct information about it which shows that with the use of artificial neural network the system can yield accurate results (see Figures 7 to 12).

For further improvement of this study there are several things that needs to be considered. One of these things is the system's database. Since in this study, the created database was only limited to a certain number of indigenous plants and trees in the Philippines, the future researcher may add more kinds of plants and trees and more

information about it in order to make its range wider. Another thing is the number of neurons used in the artificial neural network since as the database becomes wider; the number of neurons must also be increased with proper training in order to yield accurate results. When it comes to the training of neurons, using different images of the same kind of leaf helps a lot.

5. REFERENCES

Ahmed, N.. 2011, December 30. Medicinal Plants and Their Importance as Alternative Medicine. Retrieved from http://mdnasirahmed.wordpress.com/2011/12/30/medicinal-plants-their importance-as-alternative-medicine/

Bishop, C. M. 2006. Pattern Recognition. *Springer*, vii.

Chaki, J., and R. Parekh. 2011. Plant Leaf Recognition using Shape based Features and Neural Network classifiers. *International Journal of Advanced Computer Science and Applications, 2*(10), 41-47.

Gang Wu, S., F. Sheng Bao, E. You Xu, Y.X. Wang, Y.F. Chang and Q.L. Xiang. 2007. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. Institute of Applied Chemistry, Chinese Academy of Science, Department of Electronic Engineering, National Taiwan University of Science and Technology.

Gershenson, C. 2003. *Artificial Neural Networks for Beginners.* Retrieved from http://arxiv.org/ftp/cs/papers/0308/0308031.pdf

Kadir, A., L. Nugroho, A. Susanto and P. Santosa. 2011. Neural Network Application on Foliage Plant Identification. *International Journal of Computer Applications, 29*(9), 15-22.

Langner, J. 2006. *Irecog.* Retrieved from http://jens-maus.de/: http://www.jens-langner.de/

Wernick, Yang, Brankov, Yourganov, and Strother. 2010. Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine, Vol. 27, No. 4*, 25 - 38.