Image Analysis for Ethiopian Coffee Plant Diseases Identification

Abrham Debasu Mengistu  
Computing Faculty, Bahir Dar Institute of Technology,  
Bahir Dar University, Bahir Dar, Ethiopia  
P.O. BOX 26

Seffi Gebeyehu Mengistu  
Computing Faculty, Bahir Dar Institute of Technology,  
Bahir Dar University, Bahir Dar, Ethiopia  
P.O. BOX 26

Dagnachew Melesew Alemayehu  
Computing Faculty, Bahir Dar Institute of Technology,  
Bahir Dar University, Bahir Dar, Ethiopia  
P.O. BOX 26

Abstract

Diseases in coffee plants cause major production and economic losses as well as reduction in both quality and quantity of agricultural products. Now a day's coffee plant diseases detection has received increasing attention in monitoring large field of crops. Farmers experience great difficulties in switching from one disease control policy to another. The naked eye observation of experts is the traditional approach adopted in practice for detection and identification of coffee plant diseases. This paper presents an automatic identification of Ethiopian coffee plant diseases which occurs on the leaf part and also provides suitable segmentation technique regarding the identifications of the three types of Ethiopian coffee diseases. In this paper different classifiers are used to classify such as artificial neural network (ANN), k-Nearest Neighbors (KNN), Naive and a hybrid of self organizing map (SOM) and Radial basis function (RBF). We also used five different types of segmentation techniques i.e. Otsu, FCM, K-means, Gaussian distribution and the combinations of K-means and Gaussian distribution. We conduct an experiment for each segmentation technique to find the suitable one. In general, the overall result showed that the combined segmentation technique is better than Otsu, FCM, K-means and Gaussian distribution and the performance of the combined classifiers of RBF (Radial basis function) and SOM (Self organizing map) together with a combination of k-means and Gaussian distribution is 92.10%.

Keywords: Otsu, FCM, K-means, Gaussian Distribution.

1. INTRODUCTION

Ethiopian economy is highly dependent on agriculture. Among the agricultural production, coffee sub-sector plays a major role in the economy of the country. It is the biggest source of foreign currency earning and has a major contribution to Gross domestic product. Coffee is not only one of the highly preferred international beverages, but also one of the most important trade commodities in the world next to petroleum and nowadays its use as input in some food processing industries is increasing, for instance it is used as flavoring to various pastries, ice creams, chocolate, candies [24].

Coffee plant is a plant which grows in all over the world particularly in Ethiopia. In Ethiopia agricultural sector plays a central role in the economic and social life of the nation. Around 80 to 85 % of people in Ethiopia are dependent on agriculture; among 80 to 85% about 40% of the
sector contributes from cultivation of coffee [1]. In Ethiopia, cultivation of coffee contributed about 20% of the government’s annual income. Majority of Ethiopians economy depends directly or indirectly on cultivation of coffee [3]. The coffees which are found in Ethiopia are Arabica type, In Ethiopia coffee grows in every region of the country but majority are produced in the Oromia Region (63.7%) and in the Southern Nations, Nationalities (34.4%), with lesser amounts in the Gambela Region and around the city of Dire Dawa [3]. Generally in Ethiopia much of the coffee are produced in altitudes between 1,000 and 2,000 meters. The species of coffee is endemic to Africa and a number of classes are described in West, Central and East Africa [5]. Because of coffee disease constraints and global warming factors, only two types of coffee plant are nowadays commercially grown worldwide, these are Coffee canephora (Robusta) which are grown in lowlands and Coffee arabica (Arabica) that are produced in highlands of Africa. The species of coffee arabica type originated from Ethiopia especially in the province of Kaffa. During 15th century Yemen traders distributed coffee Arabica type in all over the world. Today, there are a few rainforests in the southwest and southeast Ethiopia that produces coffee plant in a large variety of shade trees [4]. Coffee Plant disease is a disease that affects coffee plants on the leaves, stems and roots. Nowadays coffee plant diseases become critical problem and can cause significant reduction in both quality and quantity of agricultural coffee products [4].

2. LITERATURE REVIEW

Different scholars gave a suggestion to detect the plant leaf diseases using a variety of approach and implementation ways as described here:

In [6], the authors focused on cotton image that identifies the infected parts from a given cotton images. The paper has two phases in order to identify the infected part. The first phase in the research is using edge detection this help the authors to detect the border of the image after completing edge detection analysis phase is conducted finally the classification of diseases is done, using the proposed Homogeneous Pixel Counting Technique for Cotton Diseases Detection (HPCCDD) Algorithm. The target of this research work is to discover the disease affected part of cotton leaf spot by using the image processing technique.

In [7], the authors proposed a framework for detection and classification of plant leaf diseases. They also used Kmeans techniques for segmentation. For extracting the values of hue, intensity and saturation form a given RGB input images the authors converted RGB into HIS color space this helps to calculate the color of a given images. After calculating colors the authors used neural network classifier for classification of plant leaf diseases.

In [14], the authors provided software based imaging techniques to automatically detect and classify plant leaf diseases. Similarly the authors include image processing techniques starting from image acquisition to classification i.e. Image pre-processing, segmentation, features extraction and classification based on neural network.

In [15], the authors have used two classifiers i.e. spatial FCM & PNN (Fuzzy C-Means and Probabilistic neural network) on cotton plant to identify the disease in cotton plant. They have used image acquisition devices to acquire images and the images are then subjected to pre-processing and noise filtering mechanisms for a given images the authors have also use spatial FCM clustering methods for segmenting the given image.

In [16], the authors have used wheat and grape diseases based on different techniques these techniques include Otsu method, image compression, image cropping and image noise removal. The authors have used neural networks including back propagation (BP) networks, radial basis function (RBF) neural networks; generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) to diagnose wheat and grape diseases.
In [17], the authors presented an assessment on methods that used digital image processing techniques on agriculture to detect, quantify and classify plant diseases from digital images in the visible spectrum.

In [18], the authors focused on plant disease identification based on image processing approach. They extracted three groups of features i.e. color, shape and texture features. In this research they used principal component analysis (PCA) for reducing the dimensions of feature space and then neural networks including back propagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) were used as the classifiers to identify wheat diseases and grape diseases, respectively. In this research the authors focus on the two kinds of grape diseases, finally the optimal recognition results were obtained from GRNNs and PNNs.

In [19], the author used the techniques of machine vision applied to agricultural science, and it has great perspective especially in the plant protection field, which ultimately leads to crops management. The author also described a software prototype system for rice disease detection based on the infected images of various rice plants [22].

In [23], the authors used SVM and Bayes on rice diseases detection. In the work of the authors, an automated system has been developed to classify the leaf brown spot and the leaf blast diseases of rice plant based on the morphological changes of the plants. The system has been validated using 1000 test spot images of infected rice leaves collected from the field, gives 79.5% and 68.1% accuracies for Bayes’ and SVM Classifier based system respectively.

In [24], the author has shown that the application of image processing on identifications of Ethiopian coffee beans based on their growing area to classify different varieties of Ethiopian coffee based on their growing regions that are found in Ethiopia (Bale, Harar, Jimma, Limu, Sidamo and Welega) which are popular and widely planted in Ethiopia.

3. STATEMENT FOR THE PROBLEM

There are so many things that can cause different diseases to the coffee plant, which damaged big crop fields and ultimately the economy of the country is affected. If coffee plant diseases are detected on the early basis and prevented accordingly, then big losses can be avoided. So to strengthen the agricultural fields and the economy of the country, rapid and accurate detection of coffee plant diseases is needed. In Ethiopia, technologies of image analysis or computer vision have not been explored in a significant manner in the development of automation of agricultural and food industries. Particularly, identification or recognition of diseases on Ethiopian coffee plant is based on traditional ways [24]. Therefore suitable actions have to be taken to control diseases on agricultural products while reducing the use of chemicals to control the diseases. In coffee plant, there are three main diseases namely: Coffee Leaf Rust (CLR), Coffee Berry Disease (CBD), and Coffee Wilt Disease (CWD). Diagnosing and recognition of coffee plant disease is very important in order to cure and control the spreading of diseases. The method of diagnosing these plant diseases is based on the knowledge of experts. Image processing and machine learning now a day’s become the key technique for the diagnosis of various features of the plant in the areas of agriculture, because it minimize confusion and helping the expert in avoiding the abuses during diagnosis and control of coffee plant diseases especially where there is shortages of experts [5]. To this end this study answers the following research questions:

- What is the suitable segmentation technique that helps us to identify CLR, CBD and CWD?
- How to develop an automatic coffee diseases identification system based using image analysis technique
- How is the performance of the segmentation techniques and identification system?
4. EXPERIMENTATION
4.1. Materials and Tools
For acquiring images of coffee plant we used canon EOS 600d camera. When images were taken, the camera was fixed on a stand which reduces the movement of hand and capturing uniform images of coffee plant. We have used three varieties of distance i.e. 110mm, 130mm and 155mm form the coffee leaf. Finally we get better image on the distance of 130mm from the coffee leaf. To obtain uniform lightning or balanced illumination we used 100W lamp. Whenever we capture images of coffee we turn on the power of lamp so as to get minimal noises of coffee plant leaf image. The images were taken at resolution of 1632x1224 pixels and finally reduce 360 X 360 pixels because this is the standard images that can be used in image processing.

4.2. Implementation Tool
MATLAB 2013Ra on windows platform is used because MATLAB is the state of the art tool for image processing and machine learning. Therefore, for the purpose of displaying, editing, processing and analyzing and recognizing coffee diseases recognition MATLAB tool were used.

4.3. Design of Coffee Diseases Identification Framework

As shown in Fig. 1. The first step for Ethiopian coffee diseases identification framework is to take the picture as an input. Once we captured the image, image processing techniques were done, the second step in Ethiopian coffee plant diseases identification is that pre-processing of image, pre-processing image is commonly used for reducing low frequency background noise, normalize the intensity of the individual particle image, removing reflection and masking portion of image. The most commonly used pre-processing steps are as follows to reduce the pre-processing time,

FIGURE 1: Coffee Disease Recognition Process Model.
image are resized to lower resolution pixel. The image is cropped for removing extra areas. In the next steps by performing some filtering tool the noises are removed the original image color RGB image are transformed into intensity one [20, 21]. The third step in our research is segmenting image. Ethiopian coffee plant diseases image segmentation is the largest part which affects the accuracy of classification or identification steps of Ethiopian coffee plant disease recognition stages [20, 21]. Image segmentation is the key behind understanding of Ethiopian coffee leave image. There are different techniques of image segmentation, but there is no one single technique that is appropriate to all image processing applications. Therefore in our research we proposed five types of segmentation techniques these are K-means, FCM, Otsu, and Gaussian distribution and finally we combined K-means and Gaussian distribution techniques for segmenting Ethiopian Coffee plant diseases to obtain better performance and identify effective segmentation techniques. In feature extraction stage, the features of Ethiopian coffee diseases are extracted to feed into the classifiers. The feature should be measurable, highly sensitive, highly correlate, high specificity, high probability of true positive and negative response. Feature extraction is extracting representing features of a given Ethiopian coffee plant diseases image. The purpose of feature extraction is to reduce the original data set by measuring properties, or features, that distinguish between the three types of coffee plant diseases. In our case we have two groups of features these are GLCM and Color features. In Ethiopian coffee plant diseases have different color variation of each type and color analysis computed by taking HSV values. The feature set that were extracted from Ethiopian coffee leave image produces very big matrices, in order to reduce the size of matrices PCA (principal component analysis) is applied finally GA (Genetic Algorithm) is used for feature selection. The final step of Ethiopian coffee diseases recognition is the classification stage. Depending upon the extracted features of coffee diseases it classifies according to the predefined class [21]. In order to train the classifiers, a set of training of coffee plant diseases image was required, and the class label where it belongs to, 9100 coffee plant diseases image were taken from regions of Ethiopia where more coffee are produced that is Southern Nations, Nationalities, Jimma and Zegie.
As shown in figure 2, the acquired coffee plant leaf disease image was converted to gray scale, and the grayscale converted to black and white. From black and white image we applied segmentation techniques that we mentioned and sobel edge detection method to find the border of the acquired coffee leaf image in addition; this helped us to extract morphological features. In order to remove noises from the leaf i.e. dust and other small particles we applied median filtering methods. Finally we get filtered and traced image of coffee plant leaf to extract features.

5. RESULTS
We have designed experimental scenarios to test the identification performance. The performances of recognition were tested by ANN (Artificial Neural Network), KNN (Nearest Neighbor classification), Naive Bayes and a hybrid of RBF and SOM (Radial basis function and Self organizing map) using four different techniques of segmentation. In order to train the classifiers, a set of training diseased coffee image was given to the model in addition to the class label of Ethiopian coffee plant image, 9100 coffee plant diseases image were collected from the regions of Ethiopia i.e. Southern Nations, Nationalities, Jimma and Zegie from Coffee Leaf Rust (CLR), Coffee Berry Disease (CBD) and Coffee Wilt Disease (CWD). From the total of 9100 data sets, 6370 were used for model training and 2730 were used for performance testing. In our research, there were three output classes, because the coffee plant diseases type were three. The representing features of training were normalized with mean 0 and variance 1 this helps the model to converge. We carried out experiments to test the performance of our model. We used a combination of RBF and SOM, in RBF, all the training data is given to the model for training. In RBF network we used one hidden layer neurons with RBF activation functions. Then one output node is used to combine the outputs of the hidden neurons. Once the network is trained using RBF, it is very simple to differentiate the diseases. Then the output of this RBF is given to SOM because we collected images of coffee plant diseases in uncontrolled environments and this helped us to take minimal epochs for choosing the activation value and also provides higher rates of convergence.

As we have discussed in detail in the previous section, the experiments were conducted under eight scenarios, by using texture and color features separately, this helps us to get the more representing features of Ethiopian coffee plant diseases and finally combining the two feature sets and also we carried out experiments for segmentation techniques i.e. Otsu-means, FCM, Gaussian distribution and a combination of K-means and Gaussian distribution. After that we were compared the performance of classifiers ANN, KNN, Naive and combination of RBF and SOM. In general, the result showed that color features have more representing power than texture features and the classification performance of combination of SOM and RBF is by far better than ANN, KNN and Naive. As indicated in Table 1-Table 5, the summary result of KNN, ANN, Naive and a combination of RBF and SOM by using Otsu segmentation are 58.9%, 52.05
%, 56.16% and 68.49% respectively. When we conducted experiments for the classifiers using K-means segmentation we have got 58.16%, 79.04%, 53.47% and 90.07% respectively. Similarly we conducted experiments for the rest of segmentation techniques. Finally we combined k-means segmentation and Gaussian distribution techniques in order to increase the performance of segmentation this is because of segmentation is a key techniques in image analysis, when we conducted experimentation using the combined segmentation techniques together with the classifiers we have got 68.49%, 80.82%, 67.12% and 92.10% respectively.

<table>
<thead>
<tr>
<th>TABLE 1: Summary Result of All Classifier using K-means Segmentation.</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>ANN</td>
<td>Naïve</td>
<td>SOM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLR</td>
<td>CBD</td>
<td>CWD</td>
<td>CLR</td>
<td>CBD</td>
<td>CWD</td>
<td>CLR</td>
</tr>
<tr>
<td>Total 2730</td>
<td>Total 2730</td>
<td>Total 2730</td>
<td>Total 2730</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correct 1588</td>
<td>correct 2158</td>
<td>correct 1460</td>
<td>correct 2459</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not correct 1142</td>
<td>not correct 572</td>
<td>not correct 1270</td>
<td>not correct 271</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% 58.16</td>
<td>% 56.164384</td>
<td>% 53.479853</td>
<td>% 68.493151</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>TABLE 2: Summary Result of All Classifier using Otsu Segmentation.</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>ANN</td>
<td>Naïve</td>
<td>SOM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLR</td>
<td>CBD</td>
<td>CWD</td>
<td>CLR</td>
<td>CBD</td>
<td>CWD</td>
<td>CLR</td>
</tr>
<tr>
<td>Total 73</td>
<td>Total 73</td>
<td>Total 73</td>
<td>Total 73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correct 43</td>
<td>correct 38</td>
<td>correct 41</td>
<td>correct 50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not correct 30</td>
<td>not correct 35</td>
<td>not correct 32</td>
<td>not correct 23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% 58.804111</td>
<td>% 52.054795</td>
<td>% 56.164384</td>
<td>% 68.493151</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 4: Performance of Coffee Diseases Identification using Otsu Segmentation.

<table>
<thead>
<tr>
<th></th>
<th>KNN</th>
<th>ANI</th>
<th>Naive</th>
<th>SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>18</td>
<td>2</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>CBD</td>
<td>6</td>
<td>16</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>CWD</td>
<td>2</td>
<td>1</td>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>

Total 73 correct 53 not correct 20
%
72.60274

FIGURE 5: Performance of Coffee Diseases Identification using FCM Segmentation.

<table>
<thead>
<tr>
<th></th>
<th>KNN</th>
<th>ANI</th>
<th>Naive</th>
<th>SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>15</td>
<td>4</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>CBD</td>
<td>7</td>
<td>13</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>CWD</td>
<td>7</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>

Total 73 correct 40 not correct 33
%
54.794521

TABLE 3: Summary Result of All Classifier using FCM Segmentation.

<table>
<thead>
<tr>
<th></th>
<th>KNN</th>
<th>ANI</th>
<th>Naive</th>
<th>SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CBD</td>
<td>14</td>
<td>14</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CWD</td>
<td>7</td>
<td>2</td>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

Total 73 correct 41 not correct 32
%
56.164384

TABLE 4: Summary Result of All Classifier using Gaussian Distribution Segmentation.

<table>
<thead>
<tr>
<th></th>
<th>KNN</th>
<th>ANI</th>
<th>Naive</th>
<th>SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR</td>
<td>19</td>
<td>3</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>CBD</td>
<td>6</td>
<td>16</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>CWD</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>6</td>
</tr>
</tbody>
</table>

Total 73 correct 52 not correct 21
%
71.232877

International Journal of Biometrics and Bioinformatics (IJBB), Volume (10) : Issue (1) : 2016
6. CONCLUSION AND FUTURE WORK

In this research paper, we have evaluated four types of classifiers (ANN, KNN, Naïve and combination of RBF and SOM) together with five different segmentation techniques for Ethiopian coffee plant diseases identification. In our Experimental simulation, the combination of RBF and SOM with a combined segmentation techniques has a better performance than the other classifiers and also the combination of K-means and Gaussian distribution has a better
performance than Otsu, K-means, FCM and Gaussian. But when we see the training time of the combination of RBF and SOM, it takes longer time in training. In addition to this, we recommend for further research and improvements on identification of Ethiopian Coffee diseases by exploring more segmentation techniques and more features on stem parts of coffee plant.

7. ACKNOWLEDGMENTS
We gratefully acknowledge Bahir Dar University, Bahir Dar institute of Technology for funding, generous assistance and valuable information provided to us by our institute. We must, however, express special thanks to Dr. Mekete Dessie, Dean of post graduate and research school of Bahir Dar institute of Technology who nevertheless have a common gift for dialogue and hospitality.

8. REFERENCES


