

An Automatic Coffee Plant Diseases Identification Using Hybrid Approaches of Image Processing and Decision Tree

Abrham Debasu Mengistu, Seffi Gebeyehu Mengistu, Dagnachew Melesew
Bahir Dar Institute of Technology, Faculty of Computing, Bahir Dar University, Ethiopia

Article Info

Article history:

Received Jan 3, 2018

Revised Feb 15, 2018

Accepted Feb 22, 2018

Keywords:

BPNN

CLR

CBD

CWD

Decision tree

ABSTRACT

Coffee Leaf Rust (CLR), Coffee Berry Disease (CBD) and Coffee Wilt Disease (CWD) are the three main diseases that attack coffee plants. This paper presents the identification of these types diseases using hybrid approaches of image processing and decision tree. The images are taken from Southern Ethiopia, Jimma and Zegie. In this paper backpropagation artificial neural network (BPNN) and decision tree had been used as techniques; a total of 9100 images were collected. From these, 70% are used for training and the remaining 30% are used for testing. In general, 94.5% accuracy achieved when decision tree and BPNN with tanh activation function are combined.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Abrham Debasu Mengistu,
Bahir Dar Institute of Technology,
Faculty of Computing,
Bahir Dar University, Ethiopia.
Email: abrhamd@bdu.edu.et

1. INTRODUCTION

Coffee plant is a plant which is grown 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 in agriculture; among 80 to 85% about 40% of the sector contributes from cultivation of coffee [1]. 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 coffee is endemic to Africa and a number of classes are described in West, Central and East Africa [2]. 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 is a few rainforests in the southwest and southeast Ethiopia that produces a coffee plant in a large variety of shade trees [3]. Coffee Plant disease is a disease that affects coffee plants on the leaves, stems and roots parts nowadays coffee plant diseases becomes critical problem and can cause significant reduction in both quality and quantity of agricultural coffee products.

2. LITERATURE REVIEW

P.Revathi & M.Hemalatha [4], in this research paper the authors focus on cotton image that identifies the infected parts from a given cotton images. Besides P.Revathi and M.Hemalatha, the identification has two phases. 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. The second phase is the classification of diseases using the proposed Homogeneous Pixel Counting Technique for Cotton Diseases Detection (HPCCDD) Algorithm.

Dheeb Al Bashish et al [5], in their work the authors have 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 are converted RGB into HIS color space this helps to calculate the color of a given images.

Prakash M. Mankar & Shreekanth Ghorpade [6], in this research, the authors provide software based on imaging techniques to automatically detect and classify plant leaf diseases. Similarly the authors include image processing techniques starting from image acquisition to classification.

Premalatha.V etal [7], in this paper, the authors have used spatial FCM & PNN (Fuzzy C-Means and Probabilistic neural network) on cotton plant to identify the disease in cotton plant. The authors 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.

Nikita Rishi & Jagbir Singh Gill [8], 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 for classification they 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 research [9], the authors' presents an assessment on methods that indicates the use digital image processing techniques on agriculture to detect quantify and classify plant diseases from digital images in the visible spectrum.

Haiguang Wang, etal [10], in their work plant disease identification based on image processing approach the authors extracted three groups of features i.e. color, shape and texture features and they used principal component analysis (PCA) for reducing the dimensions of feature space and then neural networks including backpropagation (BP) network were used as the classifiers to identify wheat diseases and grape diseases, respectively.

In research [11], the author used the techniques of machine vision that are applied to applied to agricultural science and it has great perspective especially in the plant protection field, which ultimately leads to crops management.

S. Phadikar, etal [12] in this research paper 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.

Habtamu Minasie [13], in this research paper the author shown that the application of image processing on identifications of Ethiopian coffee beans based on their growing area in view of this the authors 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.

Abrham Debasu etal [14], in their work entitled as "Ethiopian Coffee Plant Diseases Recognition Based on Imaging and machine learning" the authors have shown that the application of image processing and machine learning to identify coffee leaf diseases besides, the authors have used the combined approaches of SOM and RBF for the identification of Ethiopian coffee plant diseases.

In research [15], the authors showed that Object Detection using Haar Cascade Classifier widely applied in several devices and applications as a medium of interaction between human and computer.

In research [16], the author focuses on wood characterization that includes: hardness, strength, cleavage resistance, etc. Among these properties there that can be measured or estimated by visual observation on cross-sectional areas of wood. Edge detection is applied to the wood test images with the aim to improving the characteristics of wood fibers so as to make it easier to distinguish their quality. The authors used Naïve Bayes classifier for classification.

In research [17], the researchers showed that the rice variety and the quality of rice lead to originality certification of rice by existing institutions. The authors developed a system used as a tool to identify rice varieties. Identification process was performed by analyzing rice images using image processing. The analyzed features for identification consisted of six color features, four morphological features, and two texture features. Classifier used LVQ neural network algorithm. Identification results using

a combination of all features gave average accuracy of 70,3% with the highest classification accuracy level of 96,6% for Mentik Wangi and the lowest classification accuracy of 30% for Cilosari.

In research [18], in their work entitled as “Cotton Pests and Diseases Detection Based on Image Processing” the authors have shown that three different color models for extracting the damaged image from cotton leaf images were implemented, namely RGB color model, HSI color model, and YCbCr color model. The ratio of damage (γ) was chosen as feature to measure the degree of damage which caused by diseases or pests. This paper also shows the comparison of the results obtained by the implementing in different color models, the comparison of results shows good accuracy in both color models and YCbCr color space is considered as the best color model for extracting the damaged images.

3. RESEARCH METHODS

To collect the data set of coffee plant diseases image canon EOS 600d camera are used. When images were taken, the camera was fixed on a stand which reduces the movement of hand and capturing uniform images of coffee plant. To obtain uniform lightning or balanced illumination 100W lamp is used. Having such types of data set, it was very helpful to identify the diseases type. A total of 9100 images are considered for this study. Once the data set collected, various processing steps are performed to achieve the goal of the study through MATLAB, 2013.

The knowledge base is a central part component of the expert system for which information was obtained from an expert. Developing a knowledge base with the help of an expert as a trusted source of information is the most important thing in the expert system so that the result will be correct and valid. In this case, direct interviews with some expert of the coffee plant are conducted.

4. COFFEE PLANT DISEASES IDENTIFICATION DESIGN

The coffee plant identification system contains two basic parts the first part is building a knowledge base system and the second part is image processing part. In developing a knowledge base system extracting the knowledge of an expert and develop a rule using decision tree methods in this case, for every symptom of coffee plant diseases we should continue to apply the rule until no rules that can be applied or objective has been achieved. In case of image processing, the first stage is coffee plant diseases are given as input to the system. The second steps for coffee plant diseases recognition is that pre-processing of image, pre-processing image commonly used removing low frequency background noise, normalize the intensity of the individual particles on a given image, removing reflection and masking portion of image these is because of noises cause inaccuracy in identification of coffee plant diseases. Median filtering is used for reducing noises on coffee plant images. Image segmentation is the major techniques behind understanding of coffee plant diseases identification. There are different techniques of image segmentation, but there is no one single technique that is appropriate to all image processing applications. Therefore in this research K-means segmentation techniques are used. In feature extraction stage, the features of coffee plant diseases are extracted to feed into the classifiers. 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 three groups of features these are GLCM, Statistical and Color features. In Ethiopian coffee plant diseases they have different color variation of each type and color analysis computed by taking HSV values.

$$I = \frac{1}{3(R + G + B)}$$

$$S = 1 - \frac{\min(R, G, B)}{(R + G + B)}$$

$$H = \arccos\left\{\frac{[(R - G) + (R - B)]/2}{[(R - G)^2 + (R - B)(G - B)]^{1/2}}\right\}$$

The final step of coffee plant leaf diseases recognition is the classification stage. A classifier classifies the given datasets into their corresponding class. 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 coffees are produced that is Southern Nations, Nationalities, Jimma and Zegie.

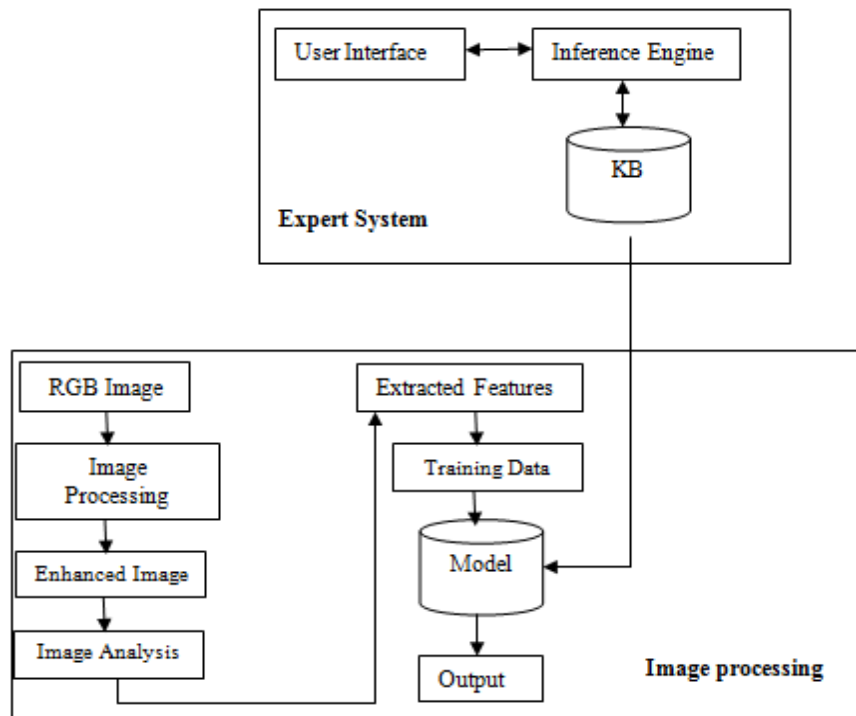


Figure1. Coffee Plant Disease Identification Model

Explaining research chronological, including research design, research procedure (in the form of algorithms, Pseudocode or other), how to test and data acquisition [1], [3]. The description of the course of research should be supported references, so the explanation can be accepted scientifically [2], [4].

5. RESULTS

We have designed experimental scenarios to test the identification performance by taking the extracted features of the diseased image. We have 17 features which are extracted from a given coffee plant image these are five GLCM, six statistical and six color features. The performances of recognition were tested by BPNN (Back Propagation Neural Network). 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. From the total of 9100 data sets, 6370 were used for model training and 2730 were used for performance testing. In this research, there are 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. The neural network needs 17 inputs of the combined feature vectors of GLCM, COLOR and STATISTICAL and 3 neurons in its output layer to classify the type. The hidden layer has 17 neurons this number was picked by trial and error methods, if the network has trouble of learning capabilities, and then neurons can be added to this layer. There is a significant change when we increase the number of hidden layers neurons until 17 but there is no change when the number of hidden layer neurons increases above 17. Each value from the input layer is duplicated and sent to all of the hidden nodes. The result indicated that there was 94.5% using BPNN with tanh sigmoid activation function. After conducting the above experiments 94.5% success achieved when backpropagation artificial neural network with tanh activation function are combined.

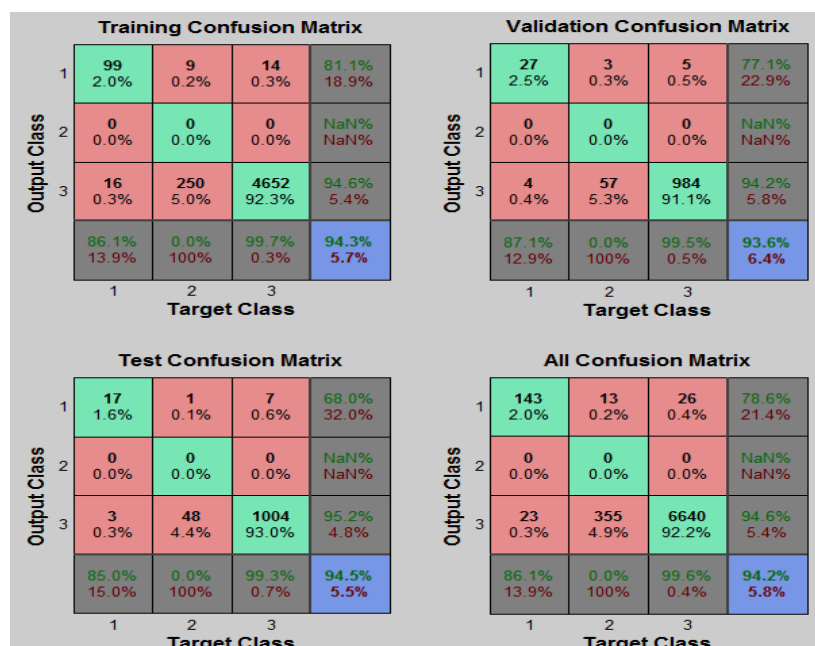


Figure 2. Confusion matrix

6. CONCLUSION

The aim of the research paper is to develop a hybrid system using decision tree and image processing techniques. In addition, this research has been focused on coffee plant diseases identification since it matches with the original purpose of the research. In this paper, BPNN with the combined features of GLCM, COLOR, STATISTICAL and knowledge base system (KB) in coffee diseases identification are tested and the accuracy of the system are presented, and the results of BPNN with tanh activation function approaches were discussed and encouraging results were obtained.

REFERENCES

- [1] Alemayehu Asfaw Amamo, "Coffee Production and Marketing in Ethiopia", *European Journal of Business and Management*, Vol.6, No.37, 2014.
- [2] Bart Minten, Seneshaw Tamru, Tadesse Kuma, and Yaw Nyarko, "Structure and Performance of Ethiopia's Coffee Export Sector", Ethiopia Strategy Support Program, 2012.
- [3] Abrham Deabsu, Dagnachew Melesew and Seffi Gebeyehu, "Image Analysis for Ethiopian coffee plant diseases identification, *CSC-journal*, Vol.9, No.4 (2016).
- [4] P. Revathi M. Hemalatha, "Homogenous Segmentation based Edge Detection Techniques for Proficient Identification of the Cotton Leaf Spot Diseases", a, *International Journal of Computer Applications*, Volume 47–No.2, June 2012.
- [5] Dheeb Al Bashish, Malik Braik, and Sulieman Bani Ahmad, "A Framework for Detection and Classification of Plant Leaf and Stem Diseases", 2010 International Conference on Signal and Image Processing.
- [6] Prakash M. Mainkar, Shreekant Ghorpade, Mayur Adawadkar, Coffee diseases Disease Detection and Classification Using Image Processing Techniques, Volume 2, Issue 4, 2015.
- [7] Premalatha.V, Valarmathy.S, Sumithra.M.G, "Disease Identification in Cotton Plants Using Spatial FCM & PNN Classifier", *IJIRCCE*, Vol. 3, Issue 4, April 2015.
- [8] Nikita Rishi, Jagbir Singh Gill, "An Overview on Detection and Classification of Plant Diseases in Image Processing", *International Journal of Scientific Engineering and Research (IJSER)* Volume 3 Issue 5, May 2015.
- [9] Jayme Garcia Arnal Barbedo, Digital image processing techniques for detecting, quantifying and classifying plant diseases, SpringerPlus, December 2013.
- [10] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, Image Recognition of Plant Diseases Based on Backpropagation Networks : *IEEE*, 2015.

- [11] Anup Vibhute, Applications of Image Processing in Agriculture: A Survey, *International Journal of Computer Applications* (0975 – 8887) Volume 52– No.2, August 2012.
- [12] Antanu Phadikar and Jaya Sil, “Rice disease identification using pattern recognition techniques”, Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), 25-27 December, 2008, Khulna.
- [13] Habtamu Minasie “*Image analysis for Ethiopian coffee classification*”, 2008, Addis Ababa University
- [14] Abrham Deabsu, Dagnachew Melesew and Seffi Gebeyehu, “Ethiopian Coffee Plant Diseases Recognition Based on Imaging and Machine Learning Techniques, *International Journal of Database Theory and Application*, Vol.9, No.4 (2016), pp.79-88.
- [15] Fitri Utaminigrum, Renaldi Primaswara Praetya, Yuita Arum Sari, ” Image Processing for Rapidly Eye Detection based on Robust Haar Sliding Window”, *IAES*, Vol 7, No 2.
- [16] Achmad Fahrurrozi1, Sarifuddin Madenda1, Ernastuti1, Djati Kerami, “Wood Classification Based on Edge Detections and Texture Features Selection”, *IAES*, Vol.6, No.5, October 2016, pp. 2167~2175.
- [17] Lilik Sumaryanti, Aina Musdholifah, Sri Hartati, “Digital Image Based Identification of Rice Variety Using Image Processing and Neural Network”, *TELKOMNIKA Indonesian Journal of Electrical Engineering* Vol 16 No 1, 2015.
- [18] Qinghai He, Benxue Ma, Duanyang Qu, Qiang Zhang, Xinmin Hou, Jing Zhao , “Cotton Pests and Diseases Detection Based on Image Processing”, *TELKOMNIKA Indonesian Journal of Electrical Engineering* Vol 11 No 6, 2013.