International Journal of Database Theory and Application
Vol.9, No.4 (2016), pp.79-88
http://dx.doi.org/10.14257/ijdta.2016.9.4.07

Ethiopian Coffee Plant Diseases Recognition Based on Imaging and Machine Learning Techniques

Abrham Debasu Mengistu[1*], Dagnachew Melesew Alemayehu[2] and Seffi Gebeyehu Mengistu[3]

[1]*Bahir Dar University, Bahir Dar Institute of Technology, Bahir Dar, Ethiopia*
Email: abrhamd@bdu.edu.et, dagnachewm@bdu.edu.et, seffig@bdu.edu.et

Abstract

Coffee plant is a plant whose seeds called coffee beans are grown in all over the world particularly in Ethiopia. The research focuses on three major type of coffee disease which occurs on the leave part of a coffee plant, these are Coffee Leaf Rust (CLR), Coffee Berry Disease (CBD), and Coffee Wilt Disease (CWD). The aim of this paper is recognition of the three types of coffee disease using imaging and machine learning techniques. The image of Coffee plant diseases were taken from the regions of Ethiopia where more coffee is produced i.e. Southern Nations, Nationalities, and Peoples, Jimma and Zegie. In this paper artificial neural network (ANN), k-Nearest Neighbours (KNN), Naïve and a hybrid of self organizing map (SOM) and Radial basis function (RBF) are used. We conduct experiment for each group of feature set in order to get a highly correlated and the more representing features. The total number of data sets is 9100. From the total of 9100, 70% were used for training and the remaining 30% were used for testing. . In general, the overall result showed that color features represents more than texture features regarding recognition of coffee plant diseases and the performance of combination of RBF (Radial basis function) and SOM (Self organizing map) is 90.07%.

***Keywords**: SOM, RBF, KNN, ANN, Coffee Arabica*

1. Introduction

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 [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 are a few rainforests in the southwest and southeast Ethiopia that produces coffee plant in a large variety of shade trees [2]. Coffee Plant disease is a disease that affects coffee plants on the leaves, stems and roots. Nowadays

ISSN: 2005-4270 IJDTA

coffee plant diseases become critical problem and can cause significant reduction in both quality and quantity of agricultural coffee products [1].

2. Literature Review

The economy of Ethiopia is highly dependent on agricultural production especially in coffee production. In Ethiopia coffee production plays a very important role in the economy because coffee production takes the first place in earning foreign currency and has a main contribution to Gross Domestic Product (GDP) of Ethiopia [18].

[P. Revathi, M. Hemalatha] [1], in this research the authors focus 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 [5].

[Dheeb Al Bashish, *et al*], in this research the authors have proposed a framework for detection and classification of plant leaf diseases they also used K means 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. After calculating colors the authors used neural network classifier for classification of plant leaf diseases [6]. Elham Omrani, *et al*. [4] in this research the authors have proposed RBF(radial basis function) for apple disease detection [7].

[Prakash M. Mainkar, Shreekant Ghorpade]. [8], 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 *i.e*. Image pre-processing, segmentation, features extraction and classification based on neural network.

[Premalatha. V, Valarmathy. S, Sumithra. M. G]. [9], in this paper, 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. 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]. [10], in this research, 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.

[Jayme Garcia, Arnal Balbedo], in this research the authors presents a assessment on methods that indicates use digital image processing techniques on agriculture to detect, quantify and classify plant diseases from digital images in the visible spectrum. [11].

[Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li], in this research the authors focus on plant disease identification based on image processing approach. The authors 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 [12].

[Phadikar. S], in this research paper 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 authors also describe a software prototype system for rice disease detection based on the infected images of various rice plants [16].

[S. Phadikar, J. Sil, and A. K. Das], 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. 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 [17].

[Habtamu Minasie], 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 author 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 [18].

Table 1. Literature Review

Paper	Features Extraction	Feature Extraction Technique	Classifier	Plant
V. A. Gulhane, *et al*[1]	1. color 2. shape of holes	Unsupervised self-organizing feature map.	ANN	Cotton
S. Arivazhagan, *et al*[2]	1. Color 2. Texture	Color-Co-Occurrence methodology (CCM).	SVM.	Native plant species
Dheeb Al Bashish, *et al*	1. Color 2. Texture	CCM	ANN	Five diseases which effect on the plants; they are; Early scorch, Cottony mold, Ashen mold, late scorch, tiny whiteness.
Elham Omrani, *et al*	1. Color 2. shape 3. Texture	Wavelets transform, SGDM and CCM.	support vector regression (SVR)	Apple
Haiguang Wang, *et al*	1. Color 2. Texture 3. Shape	Principal component analysis (PCA)	BPNN, RBFNN, GRNN and PNN	Wheat and grape diseases.
Santanu Phadikar, *et al*	1. Boundary 2. Edge 3. Spot	Boundary detection algorithm	ANN	Rice
P. Revathi, *et al*	1. Edge 2. CMYK color 3. Texture	Enhanced PSO and GA algorithm.	SVM, BPN, Fuzzy	Cotton
Habtamu Minasie	1.Color 2. Morphology	PCA	ANN	Coffee beans identification based on growing area

3. Statement of the Problem

Even if Agriculture is the leading sector of the economy in Ethiopia, the land is cultivated by old production techniques which lead to different structural problems. In agriculture disease recognition is a major challenge especially in coffee plant. 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 [7]. To this end this study answers the following research questions:

- What are the key feature that identifies CLR, CBD and CWD?
- How to develop an automatic coffee diseases recognition system based on imaging and machine learning approach?
- How is the performance of the recognition system?

4. 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.

5. Implementation Tool

For image processing of Coffee diseases recognition 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.

6. Design of Coffee Diseases Recognition

The first stage in coffee diseases recognition is input image. Coffee plant diseases are given as input to the system. The second step for coffee plant diseases recognition is that pre-processing of image, commonly used for removing low frequency background noise, normalize the intensity of the individual particles on a given image, removing reflection and masking portion of image this is because noises cause inaccuracy in identification of Ethiopian coffee plant diseases. The noises which appear on a coffee plant image are reduced by filtering method. The methods that we used for reducing noises on coffee plant image are median filtering. In feature extraction stage, the features of coffee plant diseases are extracted to feed into the classifiers. The feature should be measurable, highly sensitive, highly correlative, high specificity, high probability of true positive and negative response. 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 coffee are produced that is Southern Nations, Nationalities, Jimma and Zegie.

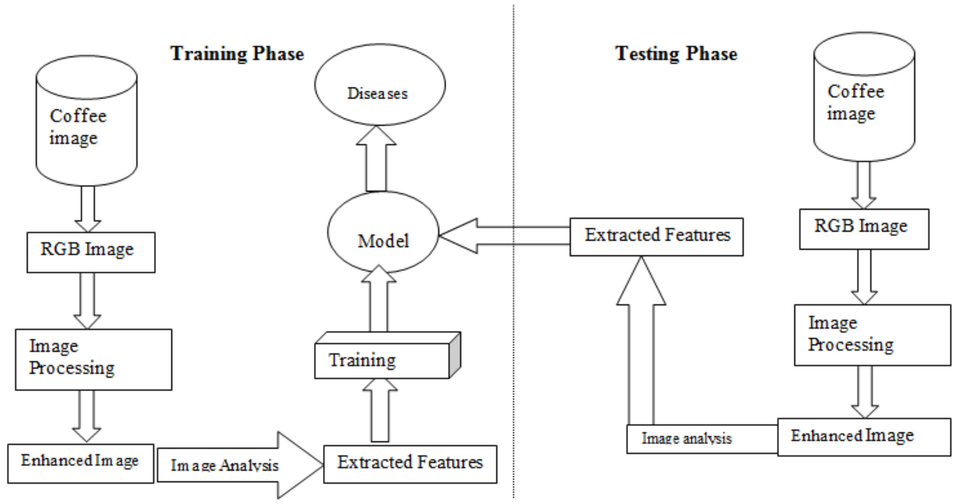
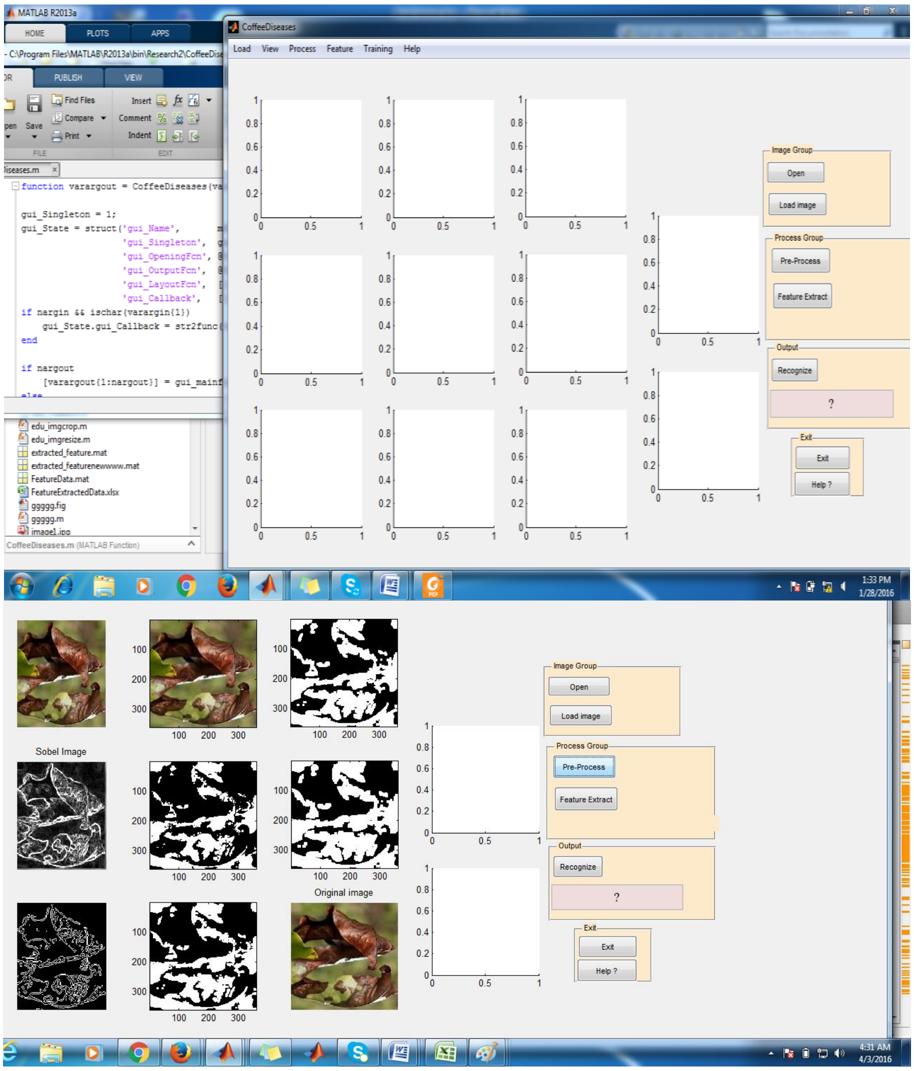


Figure 1. Coffee Disease Recognition Process Model

In machine learning and pattern recognition two fundamental phases training and testing phases are used. In the training phase, data is repeatedly given to the classifier, in order to obtain a trained model. In testing phase, the data are given to the trained model but the data are new and which are not given before these help us to know the performance of the trained model. Therefore, we need to design the model of classifier by dividing the total data set into training and testing data set. From the total of coffee plant disease image 70% was used for training and the other remaining 30% are used for testing data. As shown in Figure 1. The first step of Ethiopian coffee diseases recognition is taking the picture as an input. Once we captured the image, image processing techniques were done, the second steps Ethiopian coffee plant diseases recognition is that pre-processing of image, as we describe in the previous section, pre-processing image commonly used 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, 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 [14-15]. The third step in our research *i.e.* Ethiopian coffee plant diseases recognition is segmenting image. Ethiopian coffee plant diseases image segmentation is the largest part which affects the accuracy of the next steps of Ethiopian coffee plant disease recognition stages [14-15]. 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 selected K means techniques for segmenting Ethiopian Coffee plant diseases to obtain better performance and efficiency.

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 correlative, 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 [15].



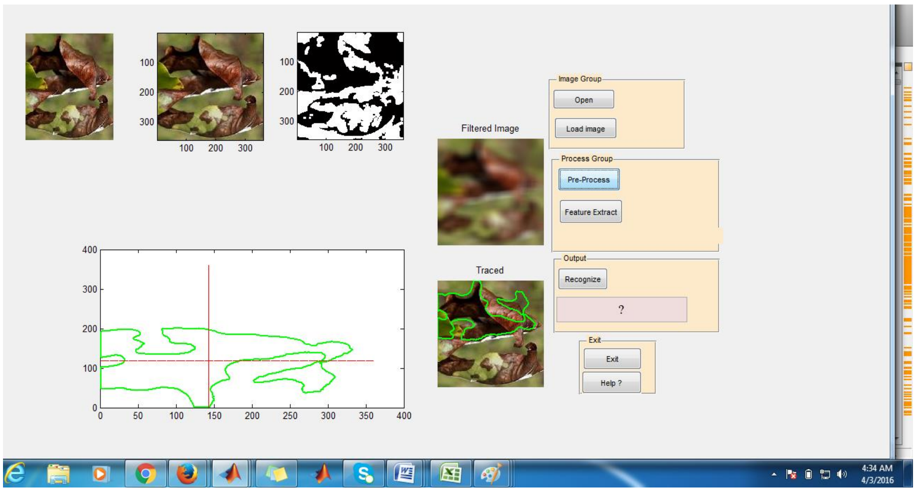


Figure 2. Coffee Diseases Recognition Prototype

As shown in the above 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 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.

7. Results

We have designed experimental scenarios to test the recognition performance by taking the extracted features of Ethiopian coffee leaf image. We have got 11 features which are extracted from a given coffee plant image these are five GLCM and six color features. 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). 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.

Table 2. Summary Result of All Classifier

		KNN				ANN	
	CLR	CBD	CWD		CLR	CBD	CWD
CLR	530	128	252	CLR	701	95	114
CBD	166	578	166	CBD	97	733	80
CWD	289	141	480	CWD	131	55	724
	Total	2730			Total	2730	
	correct	1588			correct	2158	
	not correc	1142			not correct	572	
	%	58.168498			%	79.047619	

			Naïv e			SOM	
	CLR	CBD	CWD		CLR	CBD	CWD
CLR	509	123	278	CLR	823	69	18
CBD	254	514	142	CBD	23	873	14
CWD	297	176	437	CWD	62	85	763
	Total	2730			Total	2730	
	correct	1460			correct	2459	
	not correct	1270			not correct	271	
	%	53.479853			%	90.07326	

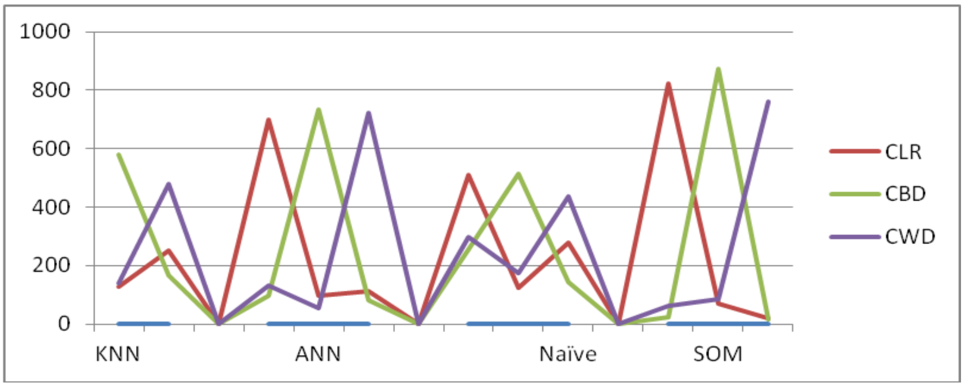


Figure 3. Performance of Coffee Diseases Recognition

As we have discussed in detail in the previous section, the experiments were conducted under three 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. After that we were compared the performance of classifiers ANN, KNN, Naïve 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 Naïve. As indicated in Table 1, the summary result of KNN, ANN, Naïve and a combination of RBF and SOM are 58.16 % , 79.04%, 53.47%, 90.07% respectively.

8. Conclusion

In this research paper we have evaluated four types of classifiers (ANN, KNN, Naïve and combination of RBF and SOM) for Ethiopian coffee plant diseases recognition. In line with this, in combination of RBF and SOM, RBF computes and the output is given to SOM. In our Experimental simulation the combination of RBF and SOM has a better performance than the other classifiers. 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 type by exploring more features not only on their leaf but also on their stem and roots.

Acknowledgment

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.

References

[1] S. N. Ghaiwat and P. Arora, "Detection and Classification of Coffee diseases Using Image processing Techniques", A Review, vol. 2, no. 3, **(2014)**.
[2] S. B. Dhaygude and N. P. Kumbhar, "Agricultural Coffee diseases Disease Detection Using Image Processing", vol. 2, no. 1, **(2013)**.
[3] URL: http://www.apsnet.org/edcenter/intropp/topics/Pages/PlantDiseaseDiagnosis.aspx
[4] URL: http://www.gardeningknowhow.com/plant-problems/disease/plant-leaf-spots.htm
[5] A. N. Rathod, B. Tanawal and V. Shah, "Image Processing Techniques for Detection of Leaf Disease", vol. 3, no. 11, **(2013)**.
[6] E. Omrani, B. Khoshnevisan, S. Shamshirband, H. Saboohi, N. B. Anuar and M. H. N. M. Nasir, "Potential of radial basis function based support vector regression for apple disease detection", ELSEVIER, Measurement, vol. 55, **(2014)**.
[7] P. M. Mainkar, S. Ghorpade and M. Adawadkar, "Coffee diseases Disease Detection and Classification Using Image Processing Techniques, vol. 2, no. 4, **(2015)**.
[8] V. Premalatha, S. Valarmathy and M. G. Sumithra, "Disease Identification in Cotton Plants Using Spatial FCM & PNN Classifier", vol. 3, no. 4, **(2015)**.
[9] N. Rishi1 and J. S. Gill, "An Overview on Detection and Classification of Plant Diseases in Image Processing", **(2014)**.
[10] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases", Barbedo Springer Plus, **(2013)**.
[11] H. Wang, G. Li, Z. Ma and X. Li, "Image Recognition of Plant Diseases Based on Back propagation Networks", IEEE, **(2015)**.
[12] K. Khairnar and R. Dagade, "Disease Detection and Diagnosis on Plant using Image processing", A Review, vol. 108, no. 13, **(2014)**.
[13] T. Acharya and A. K. Ray, "Image Processing Principles and Applications", Jhon Wiley, **(2005)**.
[14] W. K. Pratt, "Digital image processing, PIKS Scientific inside", John Wiley, 4th Edition, **(2007)**.
[15] S. Phadikar, "Rice disease identification using pattern recognition techniques", **(2008)**, pp. 420 - 423.
[16] S. Phadikar, J. Sil and A. K. Das, "Classification of Rice Leaf Diseases Based on morphological Changes", Breeding Science, **(2008)**, pp. 93-96.
[17] H. Minasie, "Image analysis for Ethiopian coffee classification", Addis Ababa University, **(2008)**.
[18] A. Salem, B. Samma and R. A. Salam, "Adaptation of K-Means Algorithm for Image Segmentation", CiteSeerx.
[19] A. Bhadane, S. Sharma and V. B. Nerkar, "Early Pest Identification in Agricultural Crops using Image Processing Techniques", vol. 2, **(2013)**.

Authors

Abrham Debasu Mengistu, he is born in February 04, 1985 and received his B.Sc. Degree in Computer Science from Bahir Dar University and also MSc. in Computer Science from Bahir Dar University, Faculty of Computing, BiT, Ethiopia. He has published 06 research papers in international journal. His main research interest is Image processing and Robotics. He is a life member of professional societies like MSDIWC.

Dagnachew Melesew Alemayehu, he is born in January 15, 1985 and received his B.Sc. Degree in Computer Science from Bahir Dar University and MSc. in Information Technology from Madras University, India. He has published 03 research papers in international journal. His main research interest is computer vision and Image processing.

Seffi Gebeyehu Mengistu, he is born in January 31, 1971 and received his B.Sc. Degree in Computer Science from Bahir Dar University and MSc. in Information Technology from Adama Science and Technology University, Ethiopia. He has published 01 research papers in international journal. His main research interest is Image processing and Data mining.

