Oilseed Leaf Disease Identification Using Image Processing

Rahel Bekele
HiLCoE, Software Engineering, Ethiopia
Information Network Security Agency
titarichel@gmail.com

Solomon Gizaw
HiLCoE, Ethiopia
Department of Computer Science, Addis Ababa University, Ethiopia
solohavi@gmail.com

Abstract

Oilseed is the most important food crop which is gaining increasing demand in the world market. Though the demand increases in the world market, there are challenges to its production. Diseases are among those challenges that resulted in crop yield loss. The diseases can be caused by bacteria, viruses, fungi or other factors. Diagnosis of diseases is traditionally done by eye inspection. This way of identifying diseases is less accurate, very time consuming, more laborious and can only be done in limited areas. Shortage of domain experts, especially in developing countries, is the major problem in early identification and management of these diseases. In addition, estimation of disease severity is prone to subjectivity.

In this paper, we propose a solution to identify oilseed crop diseases that requires less time, less effort and at the same time provides more accurate detection. The proposed solution uses image processing techniques for identification of the diseases and grading of disease severity. Grading of disease severity enhances the identification processes by providing the information which helps for taking appropriate measures or treatments depending on the disease severity.

A prototype is developed using Matlab. Leaf images of oilseed, i.e., soybean and sunflower, are collected. K-means clustering technique is implemented for segmenting the images. Then color and texture features are extracted. For identification of diseases, multi class Support Vector Machine (SVM) is used. After identification, severity grading of the disease is done using fuzzy logic approaches. The prototype is tested and diseases are identified with accuracy of 97.2% and the stage of the diseases is properly predicted. This result is more effective compared to other reviewed works that achieved accuracy of 95% and 93.5% for specific oilseed crop leaf diseases.

Keywords: Plant Leaf Disease; Image Processing; K-Means; Feature Extraction; Multi Class SVM; Fuzzy Logic

1. Introduction

Agriculture where more than 60% of the world's population depends up on remains the center of the world economy [1]. The dependency is even higher in developing countries, which is a matter of survival. It is the major source of food, income, and employment in addition to its major contribution to Gross Domestic Product (GDP) [2].

However, the potential for the production, productivity and market opportunities of oilseed is increasing [3, 4], although a number of factors stand against the quantity and quality of the yield. For instance, lack of adequate knowledge of farming, insect pests, plant diseases and limited number of experts are among the major challenges. Particularly, diseases are causing significant yield loss [5, 6] which result in social, ecological and economical loss to farmers [7].

Furthermore, inadequacy of domain experts remains a problem with the sector especially in developing countries. For instance, Biruk Ambachew [5], who conducted knowledge based oilseed disease diagnosis research in Ethiopia, claimed that the experts are not always available and may not be accessible to every farmer. Each agent has to serve on average more than a thousand farmers [5]. Moreover, farmers and even agricultural experts might fail to identify the diseases precisely [7, 8].

Identification of plant leaf diseases using image processing techniques will mainly reduce the confusion experts and farmers face during diagnosis of diseases, especially when uncommon diseases for specific areas breakout. It needs less effort and less time. It also enhances remote sensing of the diseases when experts are not available in the field.

Leaf carries feature information that enables to recognize the plant [25]. It is stated in [25] that each leaf of a plant has its own features, including oilseed crop leaves.

Works such as [26] have incorporated severity grading for Maple and Hydrangea leaves. As mentioned earlier, specific solution for specific crop has its own impact for precise recognition of diseases.

Though several researchers proposed techniques for leaf disease identification and provided significant result, they all have certain limitations [27]. According to Dhaware and Wanjale [27], to overcome the drawback of different techniques fusion of different techniques is a good idea. So in this paper we proposed best hybrid of segmentation, extraction and classification techniques for identification of oilseed crop leaf diseases and grading of the disease severity. The proposed solution will be more precise, requires less time and effort for diagnosing of diseases. On the other hand, grading of disease severity has high impact on determining the appropriate measure for mitigating the diseases. This will reduce massive yield losses and plays specific role in sustainable productivity of oilseed crops.

We designed a prototype for identification of oilseed leaf diseases using image processing technique. Fuzzy logic based approach is implemented for enhancing disease identification by pinpointing the stage of the disease severity. Fuzzy logic is multi-valued logic. The logic has more truth values than the classical logic which has only 'true' or 'false' (0 or 1) values. Incorporation of these values, between true and false or 0 and 1, enables to handle uncertainties. Hence, it results in better accuracy [13, 14]. "The uncertainties within image processing tasks are not always due to randomness but often due to vagueness and ambiguity. Fuzzy technique enables to manage these problems effectively" [14]. Fuzzy image processing in general provides a better solution to deal with uncertainties or vagueness and it is appropriate technique for expert knowledge representation and processing [14, 23] like oilseed disease severity grading which is prone to subjectivity. Such problems are difficult to resolve properly by only limiting the domain of analysis within the normal crisp set. However, representing of those uncertain or ambiguous data by allowing some degree of membership in a given set, in this case disease severity stages, will increase precision.

2. Related Work

Jakjoud *et al*. [10] proposed a solution for detection of tomato leaves disease. K-Nearest Neighbors (KNN) and SVM are used for classification of diseases. Fuzzy logic based decision is made for the classification in both approaches. The fuzzy interference system in this work contains if-then rules for identifying the healthy leaf from the infected one. Combined approach of SVM and KNN with fuzzy logic gave a better result than the global SVM and KNN. Overall accuracy of 98.38% and 82.01% have been achieved by Fuzzy combined KNN and Fuzzy combined SVM, respectively Though accuracy of 98.38% is very good, the proposed solution is limited to identifying the diseased one from the healthy one so it supports only two classes, i.e., it does not support classifying to specific disease types and severity grading.

The authors in [12] also proposed fuzzy logic based cotton leaf disease detection system. Gray-Level Co-Occurrence Matrix (GLCM) is used for extracting features and fuzzy logic for classification of diseases. An accuracy of 88% is achieved. Although cotton is one of oilseed crops, the proposed solution is limited to identification of cotton leaf diseases, i.e., it does not consider severity grading of the disease.

Gui and Mbaye [9] proposed identification of soybean leaf spot diseases using deep convolutional neural networks. The soybean images are downloaded from the Internet. Those images are then segmented using unsupervised fuzzy clustering algorithm. The proposed classification model is defined using Deep Convolutional Neural Network (Deep CNN). With this model accuracy of 89.84% is achieved. The researchers have also compared the proposed model with SVM and Visual Geometry Group (VGG 16). As a result, VGG 16 achieved a testing accuracy of 93.45% while SVM achieved 83.23%.

Liu *et al*. [28] proposed identification method of sunflower leaf disease based on Scale-Invariant Feature Transform (SIFT) point. In this work, sunflower leaf images are acquired by mobile phones. After preprocessing the collected images, SIFT algorithm is used to extract feature points of the images. Feature extraction of sunflower leaf scale consists of processes like extreme value extraction, accurate positioning of key points, key point direction determination and generating SIFT feature vectors. Identification of sunflower disease is held by matching of image feature vector. An average accuracy of 95% is achieved in recognition of four diseases.

A knowledge based system is proposed for diagnosis of oilseed crop diseases in [5]. The system has 'yes' or 'no' questions that represent the symptoms of different diseases. If all the symptoms are observed or all the answers given by the user are 'yes' then the system displays the type of the disease diagnosed and it suggests treatments for the identified disease if the user needs. It works based on the knowledge already fed to the system. It does not analyze the disease from the leaf image itself. In addition, the questions are prepared in human understandable format. As a result, some characteristics or features of the crop required for disease diagnosis remain unkempt.

3. The Proposed Solution

The proposed oilseed crop leaf disease identification system architecture has three major components as shown in Figure 1. These are *training*, *testing* and *grading* the stages of the disease severity for the identified disease.

3.1 Image Preprocessing

All leaf images used in this work are downloaded from reliable web sites for research on plant diseases. Domain experts from the Ethiopian Institute of Agricultural Research (EIAR) examined the validity of the downloaded images. However, each image might be captured from different distance in different environmental settings. So image preprocessing is needed to enhance the images. To come up with better results, we applied image resizing and image adjustment techniques.

We downloaded a total of 66 images which is not enough to train the model. To increase the effectiveness of the model, some techniques are applied to increase the data size like altering noises on the original image where ‘Salt and pepper’ noise is added. In addition to adding noise, flip technique is also applied to the images. As a result, a total of 3* 66 = 198 images are used as a dataset.

3.2 Image Segmentation

Before segmenting the images, color space conversion is done from RGB to L*a*b*. This conversion is required because the L*a*b* color space is device independent which means the device from which it is created or it is displayed does not affect its color definition [15, 16, 17]. In addition, this color space is more preferable in representing color [15, 16]. In L*a*b* color space, all the color information exists in a* and b* spaces while the L represents the luminosity [15, 18]. It also best fits the segmentation technique we use which is K-means.

Image segmentation is the most important part in image processing. It is performed to divide an entire image into several parts. This division is made simplify image analysis. Segmentation may depend on various features of the image like color and texture. In this paper, the segmentation technique preferred to be used is the color based k-means clustering. K-means is a partitioning technique that clusters a given data X into k mutually exclusive clusters [19]. Then it assigns an index of the cluster (idx) for each observation using Equation 1.

$$idx = kmeans(X, k) \tag{1}$$

K-means starts with assigning a centroid for each cluster. The number of clusters is pre-defined and initialization of the center of each cluster is done by squared Euclidean distance metric and the k-means++ algorithm [21, 22].

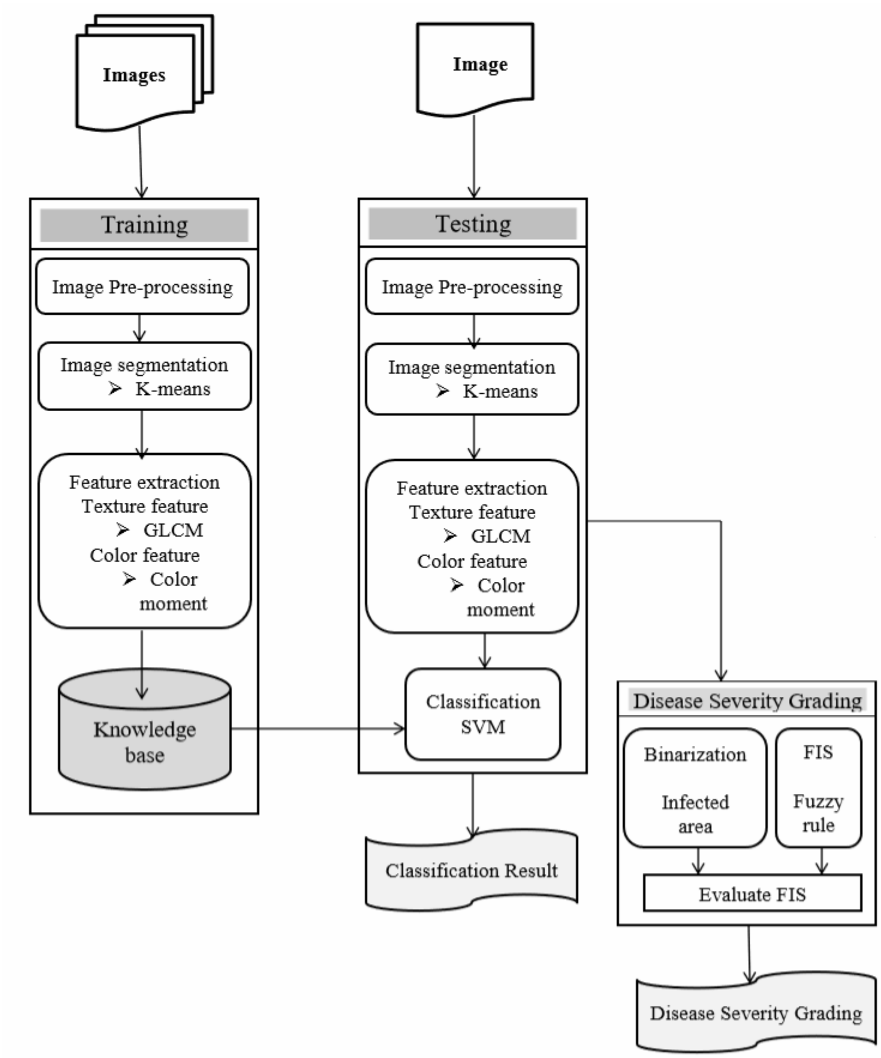

Figure 1: The Proposed System Architecture of Oilseed Crop Leaf Disease Identification

3.3 Feature Extraction

Feature extraction is required when the input image is assumed to be too large to be processed and notoriously (much data, but not much information) redundant. There are different feature extraction methods. In this paper color and texture feature extractions are considered.

a. Color Feature Extraction

Color moment is used for extracting color features. In color moments, images are assumed to be represented by a probability distribution of colors [36]. Keen [36] defines moments as the characterizations of those probability distributions. As Stricker and Orengo [24] stated, if the probability distributions of images follow certain pattern, those images are assumed to be similar. So the measured moment of the image can be used for extracting color features of an image [36]. The first three moments *Mean*, *Standard Deviation*, and *Skewness* are used by Stricker and Orengo [24]. We used Mean and Standard Deviation as feature variables.

The first color moment, Mean, can be stated as the average color value in the image whereas the second color moment, Standard Deviation, is the square root of the Variance of the distribution [36].

b. Texture Feature Extraction

There are several techniques in image processing for extracting texture features. In this work, GLCM is used.

GLCM uses graycomatrix function. This function calculates GLCM by considering spatial relationship of pixels (those pixels that are in gray-level) [30, 31, 32] as shown in Figure 2 [33]. Each element (i, j) in the resultant GLCM is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image.

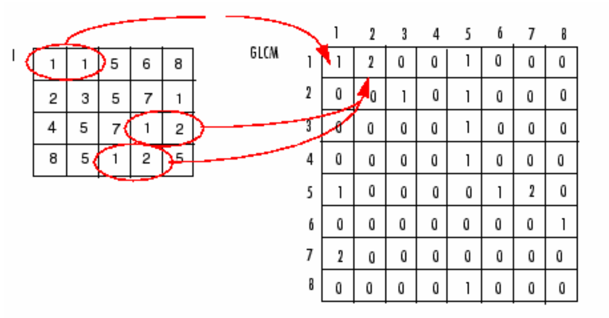

Figure 2: Process of Creating GLCM

Statistics can characterize the texture of an image because they provide information about the local variability of the intensity values of pixels in an image [33]. Statistical measures such as local standard deviation of an image and local entropy of a grayscale image are obtained from the created GLCM. The texture features are extracted using these statistical measures. In this paper using GLCM, 22 texture features are extracted. They are used for training the SVM. The selection of those texture features is based on the accuracy result they have brought during training of the model.

3.4 Image Classification and Identification

The extracted features are used as input for this step. In this step machine learning approaches are followed in many researches. Among those the supervised learning approach multi-SVM with ECOC (Error-Correcting Output Coding) algorithm is used in this paper. SVM is known for its suitability for classification when there exist few number of datasets or many attributes are considered [11, 29]. This is the major reason that this approach is selected.

Since SVM uses different kernel functions, for training of the model comparison of accuracy rate is done among kernel functions. The cubic kernel function is selected for training the model, which has provided better accuracy. This model then predicts the category/class for new data based on the dataset that it is trained with. Hence, it identifies the type of leaf disease that the plant is infected with.

3.5 Disease Severity Grading

Other than identifying oilseed crop leaf diseases, the proposed solution also provides a way for knowing in which level (stage) the disease is. For doing so fuzzy logic is applied.

In this paper oilseed severity grading Fuzzy Inference System (FIS) is designed for grading the stage of the diseases. For doing this, the percentage of area of infected region of the leaf is taken as an input for the system and the stage of the severity is given as an output.

For calculating the percentage of infected region, first both the segmented and the whole images are binarized. Then the sum of white pixels is segmented and binrazed image is divided by the sum of pixels in the whole and binarized images. Finally, the result is multiplied by hundred to get the percentage of the infected region of the leaf image.

For both input (infected area) and output (severity stage), a triangular membership function is used since it is the simplest and the most effective [34]. Triangular membership function can be represented mathematically as [35]:

$$f(x,a,b,c) = \max\left\{min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),0\right\}$$

where a and c are the “feet” of the triangle and b is the “peak”.

In our prototype the severity has six stages. The FIS predicts the stage of severity (output) based on fuzzy rules that have already been set during the design of the FIS.

4. Result

Classification learner app of Matlab is used for training and testing the model. Multi-SVM is chosen for modelling the prototype. Among quadratic, Gaussian and cubic kernel functions that the SVM uses, best accuracy is achieved while training the model with cubic kernel function.

The dataset prepared in the preceding image processing phases are used to evaluate the model. A total of 28 features which consist of 22 texture features and 6 color features are used. The total dataset consists of 198 datasets. The dataset consists of both diseased and healthy leaves of oilseed (soybean and sunflower). All these data are divided for training (training set) and testing (testing set). During training 162 (82% of the dataset) of the data are used for both training and validating and the remaining 36 test set (18% of the dataset) are used for testing. During the test almost all the test data are correctly classified to their category or class. Accuracy of 97.2% is achieved with overall error of 2.8%. Accuracy is calculated as the sum of correct classification (true positive) divided by the total number of classifications.

5. Conclusion and Future Works

With the proposed solution a very good result is achieved for identification of oilseed leaf diseases as well as disease severity grading. This will make a big difference in the experience and trends of oilseed leaf disease identification methods which is manual in most developing countries. The system has impact on reducing the time it takes and the effort required for identifying diseases. Since domain experts are few, especially in developing countries, it supports domain experts as well as farmers who may not have adequate knowledge on identifying diseases. The treatment and measures taken after a crop is identified as infected depends on the level of disease severity. On top of the identification, the proposed solution has also the capability of leveling disease severity which helps to mitigate diseases in appropriate way. This boosts the efficiency of disease identification and management experiences. This in tern prevents bigger yield loss. So it plays a significant role in sustaining the production of oilseed crops.

Severity grading of oilseed diseases works well by considering the percentage of affected regions. Future work will consider the development stage of the plant that is believed to provide better results.

References

[1] World Bank, “Agriculture and Food”, https://www.worldbank.org/en/topic/agriculture/overview, Last accessed on September 12, 2019.

[2] Neha Khanna and Praveen Solanki, “Role of agriculture in the global economy,” in Proceedings of the 2nd International Conference on Agricultural and Horticultural Sciences, Hyderabad, India, February 2014.

[3] United States Department of Agriculture, “Global Markets: Oilseeds – Consumption Grows Despite Slowing Trade, Production, “2019, https://agfax.com/2019/06/12/global-markets-oilseeds-consumption-grows-despite-slowing-trade-production, Last accessed on December 06, 2019.

[4] European Commission, “Oilseeds and Protein Crops market situation,” https://circabc.europa.eu/sd/a/215a681a-5f50-4a4b-a953-e8fc6336819c/oilseeds-market%20situation.pdf, Last accessed on December 01, 2019.

[5] Biruk Ambachew, “Knowledge Based System for Oilseed Crop Disease Diagnosis and Treatment,” Addis Ababa University, March 2015.

[6] Sachin D. Khirade and A. B. Patil,” Plant disease detection Using image processing,” in Proceedings of the IEEE International conference on computing communication control and automation, February 2015.

[7] Surender Kumar and Rupinder Kaur, “Plant Disease Detection using Image Processing - A Review”, International Journal of Computer Applications, Vol. 124, No. 16, 2015.

[8] Pallavi S. Marathe, “Plant Disease Detection using Digital Image Processing and GSM,” International Journal of Engineering Science and Computing, Vol. 7, No. 4, 2017.

[9] Jiangsheng Gui and Mor Mbaye, “Identification of Soybean Leaf Spot Diseases using Deep Convolutional Neural Networks”, International Journal of Engineering Research & Technology, Vol. 08, No. 10, 2019.

[10] F. Jakjoud, A. Hatim, and A. Bouaddi, “Detection of diseases on tomato leaves based on Sub-Classifiers Fuzzy Combination”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, No. 5, 2019.

[11] Sonal P. Patil and Rupali S. Zambre, ”Classification of Cotton Leaf Spot Disease Using Support Vector", International Journal of Engineering Research and Applications, IJERA, Vol. 4, No. 5, pp. 92-97, May 2014.

[12] Vinaya Mahajan and N. R. Dhumale, “Leaf Disease Detection Using Fuzzy Logic”, International Journal of Innovative Research in

Science, Engineering and Technology (IJIRSET), Vol. 7, No. 6, 2018.

[13] Juhi Singh, "A Parameterized Comparison of Fuzzy Logic, Neural Network and Neuro Fuzzy System: A Literature“, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 5, No. 5, 2016.

[14] Mahesh Prasanna and Shantharama Rai, "Applications of Fuzzy Logic in Image Processing – A Brief Study”, International Journal of Advanced Computer Technology, Vol. 4, No. 3, 2015.

[15] Mokrzycki W. S. and Tatol M., ”Perceptual difference in L*a*b* color space as the base for object colour identfication”, in Proceedings of the 1st International Conference on Image Processing and Communications, 2014.

[16] Vivek Singh Rathore1, Messala Sudhir Kumar, and Ashwini Verma, “Colour Based Image Segmentation Using L*a*b* Colour Space Based on Genetic Algorithm”, International Journal of Emerging Technology and Advanced Engineering, IJETAE, Vol. 2, No. 6, 2012.

[17] Dibya Jyoti Bora, “Importance of Image Enhancement Techniques in Color Image Segmentation: A Comprehensive and Comparative Study”, Indian Journal of Scientific Research, Vol. 15, pp. 115-131, 2017.

[18] Richard S. Hunter, “Photoelectric Color Difference Meter”, Journal of The Optical Society of America, Vol. 48, No. 12, pp. 985-995, 1958.

[19] Aastha Joshi and Rajneet Kaur, ”A Review: Comparative Study of Various Clustering Techniques in Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, IJARCSSE, Vol. 3, No. 3, 2013.

[20] MathWorks, “k-Means and k-Medoids Clustering,” Statistics and Machine Learning Toolbox, 2015.

[21] David Arthur and Sergei Vassilvitskii, ”k-means++: The Advantages of Careful Seeding”, in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA, New Orleans, Louisiana, 2007.

[22] Youguo Li and Haiyan Wu, “A Clustering Method Based on K-Means Algorithm”, in Proceedings of the 2012 International Conference on Solid State Devices and Materials Science, Macao, 2012.

[23] Abdallah A. Alshennawy and Ayman A. Aly, “Edge Detection in Digital Images Using Fuzzy Logic Technique”, International Journal of Computer and Information Engineering, Vol. 3, No. 3, 2009.

[24] Markus Stricker and Markus Orengo, "Similarity of Color Images”, in Proceedings of SPIE - The International Society for Optical Engineering, San Jose, 1995.

[25] Surender Kumar and Rupinder Kaur, “Plant Disease Detection using Image Processing - A Review”, International Journal of Computer Applications, Vol. 124, No. 16, 2015.

[26] S. Lokesh, D. Naveenkumar, K. Rajesh, G. A. Ramesh Kamath, and Martin Joel Rathnam, “Leaf Disease Detection and Grading using Computer Vision Technology and Fuzzy Logic”, International Journal of Innovative Research in Science, Engineering and Technology, IJIRSET, Vol. 6, No. 3, 2017.

[27] Chaitali G. Dhaware and K. H. Wanjale, "A Modern Approach for Plant Leaf Disease Classification which Depends on Leaf Image Processing", in Proceedings of the International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, January 2017.

[28] Jun Liu, Fang Lv, and Bobo Liu, "Identification Method of Sunflower Leaf Disease Based on SIFT Point", Journal of Image and Graphics, Vol. 7, No. 2, pp. 64-67, 2019.

[29] Jain S., “A Machine Learning Approach: SVM for Image Classification in CBIR”, International Journal of Application or Innovation in Engineering and Management, Vol. 2, No. 4, 2013.

[30] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, “Image Texture Feature Extraction Using GLCM Approach”, International Journal of Scientific and Research Publications, Vol. 3, No. 5, 2013.

[31] Leen-Kiat Soh and Costas Tsatsoulis, “Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices”, Journal of IEEE Transactions on Geoscience and Remote Sensing (TGRS), Vol. 37, pp. 780-795, 1999.

[32] Robert M. Haralike, K. Shanmugam, and Itshak Dinstein, “Textural features for image classification”, Journal of IEEE Transactions on Systems, Man, and Cybernetics, Vol. 3, No. 6, 1973.

[33] MathWorks, “Texture Analysis”, Image Processing Toolbox, 2015.

[34] Jin Zhao and Bimal K. Bose, “Evaluation of membership functions for fuzzy logic controlled induction motor drive”, in Proceedings of the IEEE 2002 28th Annual Conference of the Industrial Electronics Society, IECON 02, Sevilla, Spain, 2002.

[35] MathWorks, “Fuzzy Inference System Modeling”, Fuzzy Logic Toolbox, 2015.

[36] Noah Keen, “Color Moments,” 2005, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0405/KEEN/av_as2_nkeen.pdf, Last accessed on December 02, 2019.