

A supervised Multi-Spectral Image Classification for Remote Sensing Data

Akram M. Zeki

Department of Information System
International Islamic University Malaysia
Kuala Lumpur, Malaysia
akramzeki@iium.edu.my

Muhsin A. Zaid

Department of Information Systems
International Islamic University Malaysia
Kuala Lumpur, Malaysia
muhsin_shamry@yahoo.com

Abstract— With the advent of photography equipment and techniques combination to revolution of computer and digitalization in both hardware and software, it takes another dimension. This research shade some light on the Multi-Spectral Image Classification and the importance of this field in Image processing. The supervised classification approach was considered in this research where three of its types were explained, Minimum Distance (MD), Maximum Likelihood (ML), and Probabilistic Neural Network (PNN). The research involves designing a package for Multi-Spectral Image classification. This includes reading data, apply Principal Component Analysis (PCA) as a feature extraction, then apply False Colour Composite (FCC) as one of the classification techniques in multi-spectral images. The research focuses on the supervised method throughout.

Index Terms — *Minimum Distance (MD), Maximum Likelihood (ML), Probabilistic Neural Network (PNN), Principal Component Analysis (PCA), False Colour Composite (FCC).*

I. INTRODUCTION

The first Landsat Multispectral Scanner System (MSS) launched in 1972, with its 4 spectral bands, each about 100nm wide, and 80m-pixel size, began the modern era of land remote sensing from space. Remote-sensing systems now exhibit a diversity and range of performance that make the MSS specifications appear modest indeed. Remote sensing is defined as the measurement of object properties on the earth's surface using data acquired from aircraft and satellites. It is therefore an attempt to measure something at a distance, rather than in situ. Since we are not in direct contact with the object of interest, we must rely on propagated signals of some sort, for example optical, acoustical, or microwave [1].

There are essentially two broad classes of satellite program: those satellites that sit at geostationary altitudes 37,746 km above the earth's surface, usually these types of satellite are generally associated with weather and climate studies, and those which orbit much closer to the earth's surface and that are generally used for earth surface and oceanographic observations. Usually, the low earth orbiting satellites are in a sun-synchronous orbit, in that their orbital plane processes around the earth at the same rate that the sun appears to move across the earth's surface. In this manner the satellite acquires

data at about the same local time on each orbit with an error of 4 minutes, 23:56 for one orbital [2].

Supervised Classification is based on the statistics of training areas representing different ground objects selected subjectively by users on the basis of their own knowledge experience. The classification controlled by user's knowledge but, is constrained and may even be biased by their subjective view. The classification can therefore be misguided by inappropriate or inaccurate training area information and/or incomplete user knowledge [3].

This knowledge is usually obtained from one or more sources. The most common sources are topographical maps, field survey and measurements and on many occasions this knowledge is gained by the means of visual interpretation of remotely sensed imageries after being subjected to specific types of enhancement.

The remainder of this paper is organized as follows. It starts with an introduction which is followed by a literature review. The proposed methodology follows and the results analysis and finally the conclusion which is then followed by references.

II. RELATED WORK

This is based on the statistics of training areas representing different ground objects selected subjectively by users on the basis of their own knowledge experience. The classification controlled by user's knowledge but, on the other hand, is constrained and may even be biased by their subjective view. The classification can therefore be misguided by inappropriate or inaccurate training area information and/or incomplete user knowledge [3]. In other words, the classifier is directed according to some prior knowledge. This knowledge is usually obtained from one or more sources. The most common sources are topographical maps, field survey and measurements and on many occasions this knowledge is gained by the means of visual interpretation of remotely sensed imageries after being subjected to specific types of enhancement.

In the classical approach, two algorithms are the most commonly used. These are the Minimum Distance (MD) and the Maximum Likelihood (ML). Minimum Distance based on Euclidean distance measurement between each of the multi-spectral pixel 1s that in pattern recognition term, is usually referred to as pattern vectors and the centers (mean vectors) of

the classes. [4] Uses MD as one method to calculate the nearest pattern of the ANN results to the input pattern. Another paper used MD to optimize simultaneously to obtain the correct partitioning for a given number of clusters present in an image [5]. Dr Kayla used MD when the authors proposed for using Support Vector Machines (SVM) in the paper [6].

Maximum Likelihood (MLE) is based on the assumption that each class can be described by some statistical properties that encode its distribution in the multi-spectral feature space. In this type of classification a normal distribution model is usually considered, which is specified by covariance matrices and the mean vector of the classes. Maximum likelihood is, in general, more accurate since it uses statistical properties of the class distribution in its model [7]. The authors of [8] used MLE when they presented a comparative object analysis of Support Vector Classification (SVC). MLE algorithms used with Support Vector Machine (SVM) and Decision Tree (DT) in land cover mapping [9].

Artificial Neural Network (ANN) approach has been used widely for classifying Remote Sensing data and generating thematic maps for various applications. Several networks have already been used for classifying remotely sensed data in supervised mode; the most common are the ART-II by [10], the Radian Base Function (RBF) network by [11], and the Probabilistic Neural Network (PNN) by [12]. The process of classification in the (ANN) approach is usually determined by the activation functions of the given networks, which are equivalents to the decision functions of the conventional methods of classification. The publishers of [13] combined the probabilistic neural network (PNN) with the Multiscale Auto Regressive (MAR) when they proposed an effective multiscale method for the segmentation of the Synthetic Aperture Radar (SAR) images via probabilistic neural network in the paper. PPN also used as algorithm to recognition the planet leaves in order to detect planet disease by [14].

In [15], research was done Hyper Spectral Images (HSIs). The HSIs characterizes the objects of interest (e.g., land-cover classes) with unprecedented accuracy. This keeps inventories up to date. The research studies the challenges of hyperspectral image classification, whose popularity and attracted interest is increasing in the field of science such as machine learning, image processing, and computer vision [15]. Similarly, the authors of [16] addressed the problem of automatic land-cover map which updates the multi-temporal and multi-spectral remote sensed images. When a pair of image is acquired on the same geographical area at two distinct time instants, it was assumed that the trained data(s) are available for either of the acquisitions (It is called source domain image) and its able to classify other image(s) (called target domain image) for which no reliable reference map is available [16].

Another work proposed a novel approach that uses ensemble of semi-supervised classifiers in remotely sensed images for detection purpose. The work presented a multiple classifier system in semi-supervised (leaning) framework instead of using a single weak classifier [17]. In [18], the authors reviewed, organizes and give analysis and comparison method used to identify change. This significantly helps in reducing

the conceptual overlap present in existing literatures giving a succinct nomenclature with which to understand and apply change detection workflows.

A similar approach by [19] presents a hybrid fuzzy classifier for land-use/land-cover (LULC) mapping. The work follows a Bayesian method to incorporate spatial contextual information into the fuzzy noise classifier (FNC) where the FNC was selected to detect noise using spectral information more accurately than its fuzzy counterparts. Markov random fields (MRFs) was used to modelled the spatial information at the level of the second-order pixel neighborhood [19].

Reference [20] developed an innovative wetland and invasive plant mapping mechanism with three integrations: one is to integrate image interpretation with feature extraction, the second is the integration of high spatial-resolution images with high spectral-resolution images, and the third is the integration of field reference data with interpreted and classified images. The mechanism followed standard procedures while performing the integration of NAIP (National Agriculture Imagery Program) and Landsat images with multiple processes of ground truthing, image classification, and validation [20].

III. PROPOSED MECHANISM

Supervised Classification can be implemented by either of three approaches, these are: classical approach, Artificial Neural Network (ANN), and integration of both [21].

a. Minimum Distance Classifier

MD is one of the simplest yet computationally cheapest methods relatively very effective algorithms for supervised multi-spectral image classification. It is based on Euclidean Distance measurements between pattern vectors and the centers of pre-identified classes in the space of selected features. For M pre-identified classes there are z_1, z_2, \dots, z_m mean vectors each represents the center of a class.

Distances between any tested pattern vector and the centers of the classes are calculated via the following Equation:

$$D_i = \|x - z_i\| = \sqrt{(x - z_i)'(x - z_i)}, \text{ for } i=1,2,\dots,M \dots (1)$$

Where: x is the pattern to be classified and D_i is the Euclidean distance. The decision is to which class x belongs is as follows :

x belongs to class l , if and only if, $D_i \leq D_j$ for all $l \neq j$.

Equation (3.10) can be modified to Equation (3.11) according to (Tou & Gonzalez, 1974) for computation efficiency and more simple form as shown below:

$$d_i = Xz_i - 0.5z_i'z_i \text{ for } i=1,2,3,\dots,M \dots (2)$$

The decision as to which class x belongs is reversed as follows:

x belongs to class i , if and only if, $d_i \geq d_j$ for all $i \neq j$.

b. Maximum Likelihood Classifier

This method of supervised classification uses statistical properties of the classes such as the mean vector and the variance/covariance matrix. Therefore, it is supposed to be

more accurate. However, computationally it is expensive as the calculation of Variance/Covariance matrix is time consuming. The Equation of Maximum likelihood, Equation (3.12), is based on Bayes function and it is defined by (Duda et al., 2001) as follows:

$$d_i(x) = \ln P(W_i) - 0.5 \ln |C_i| - 0.5 [(x-z_i)' C_i^{-1} (x-z_i)] \dots (3)$$

Where: i refers to the number of current class and $i = 1, 2, \dots, M$
 M is the total number of classes
 $P(W_i)$ is the prior probability
 C_i is the variance covariance matrix of class i $|C_i|$ is the determinant of Matrix C_i
 x is the input vector to be classified
 z_i is the mean vector of class i .

Artificial Neural Network (ANN) approach

Artificial Neural Network (ANN) approach has been used widely for classifying Remote Sensing data and generating thematic maps for various applications. The process of classification in the (ANN) approach is usually determined by the activation functions of the given networks, which are equivalents to the decision functions of the conventional methods of classification. [22] combined the probabilistic neural network (PNN) with the Multiscale Auto Regressive (MAR) when they proposed an effective multiscale method for the segmentation of the Synthetic Aperture Radar (SAR) images via probabilistic neural network in there paper. PPN also used as algorithm to recognition the planet leaves in order to detect planet disease by [22].

a) PNN Architecture

The net consists of input layer and two other layers one is named pattern layer and the other is named output layer. The units of the pattern layer are the patterns examples used for training. Thus, the number of pattern vectors used in training equals the number of neuron in the pattern layer. Figure (1) shows the architecture of this network. According to this architecture, the pattern layer is divided into groups of units each group corresponds to one class. Each neurons corresponding to one class in the output layer where the output at each neuron is specified by summing the output of each group units.

a) PNN Training

In the training there is no need for weight initialization and weight update since the weights at the connectionists of the pattern layer are the pattern examples used in the training.

That is the patterns that correspond to each class are considered as the weights between the input neurons and the pattern units of the first class. The weights at the output units are set all to 1s.

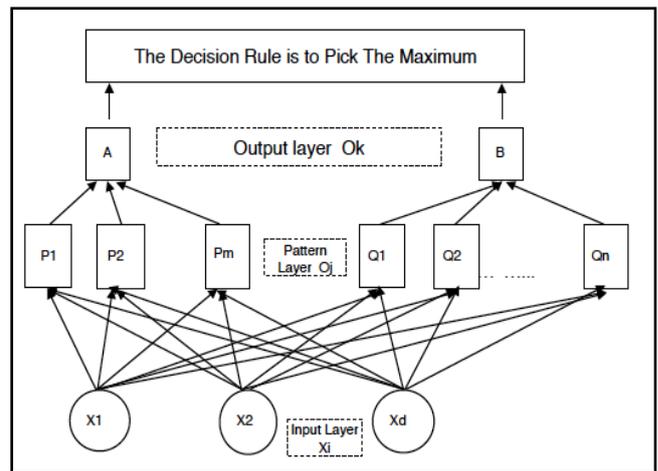


Figure 1: Probabilistic Neural Network Architecture

b) PNN Testing

In the testing phase the following procedures are done:
 Step One: The tested unit is presented to the network and after being normalized to a unit vector of length one.
 Step Two: The activation function at each unit of the pattern layer is calculated using the Equation (4):

$$O_j = \text{EXP}[(\sum W_{ji} x_i - 1)/\sigma^2] \dots (4)$$

Where: O_j is the output of the activation function at unit j of the pattern layer, W_{ji} is the weights from unit i of the input layer to pattern unit of the pattern layer, X_i is the tested pattern (pattern to be classified) and σ is the width of the Gaussian function that describe the Bayes function
 Step Three: The activation function at each unit of the output layer is calculated using the Equation (5):

$$O_q = 1/M [P(O_j) \sum w_{qj} O_j] \dots (5)$$

Where: O_q is the activation at unit q of the output unit, m is the number of units in the output layer, which corresponds to the number of classes in the output image.
 $P(O_j)$ is the prior probability factor that is used in the Bayes classifier.
 W_{qj} is the weight vector over the connectionists between the group of class unit j and the output unit q and O_j is the output at unit j of the pattern layer.
 Step Four: The decision is taken as the class corresponds to maximum output of O_q .

IV. RESULTS ANALYSIS

The first data is a multi-spectral image comes from TM-Sensor Landsat 7, comes with size (512X512), this Thematic Mapper sensor has seven bands within the visibles, NIR, MIR (two bands), and TIR, with spatial resolution of 30 meter, except band six with 120 meter resolution and due to this fact, band 6 will be excluded from research processing. The images represent Sheikh Ibrahim, an area southwest of Mosul city in Iraq. This area consists, mainly of natural vegetation at the upper part of the image and irrigated vegetation at the lower part of the image. The article of author [23] is mounting at the

middle of the image. Figure (2) shows the six original bands of Sheikh Ibrahim area (Bands 1, 2, 3, 4, 5, 7). Table (1) shows the Variance / Covariance matrix of the six original bands and the corresponding Eigen-Vector and the corresponding Eigen-Value of the PCA.

What can observe from this table that there are some bands with less variance which are (1, 4, 7). When FCC applied for rest of bands (2, 3, 5) as Red, Green, Blue (RGB) that means much of variance to be neglected, and much more of neglect in case of applying FCC with (4, 3, 1) bands as RGB. Appendix IV illustrates the FCC compensations of bands and the advantage of such compensation.

Figure (2) shows the results of bar graph on the six original images bands. Figure (4) is a bar graph shows the advantage of PCA (Data redundancy). Figure (5) shows the FCC for the original image bands (4, 3, 1) and the FCC for the PCA for the same bands. The visual inspection showed that the image was made from three PCA is better than the image was made from the original bands. Because different feature were have distinct colours where in the FCC of the original images the same features have a different shades of red colour.

The next step after applying PCA and FCC and to take the advantage of them is to apply the classification methods.

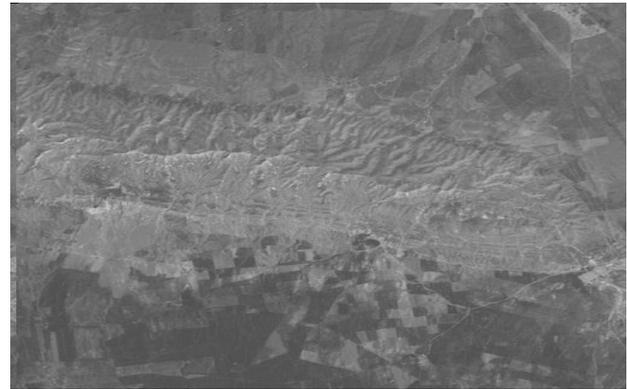


Figure 4: TM Sensor Band using PCA



Figure 5: PCA and FCC Image view

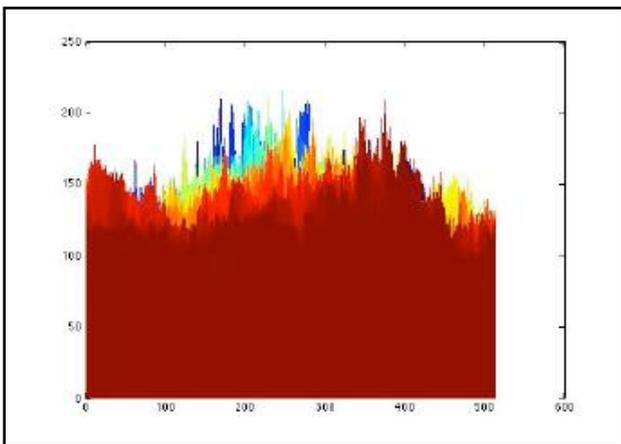


Figure 2: original Image

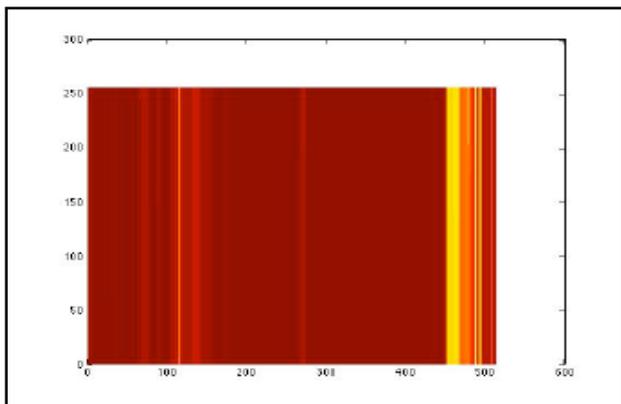


Figure 3: Bar graph for original image

Table 1: Variance / Covariance Matrix for an Image

	1	2	3	4	5	7
1	363.18	138.03	252.90	8.6574	605.81	421.38
2	138.03	160.74	277.30	36.096	253.51	172.16
3	252.90	277.30	498.28	31.406	469.15	323.75
4	8.6574	36.096	31.406	244.35	5.8934	18.897
5	605.81	253.51	469.15	5.8934	1193.2	825.81
6	421.38	172.16	323.75	18.897	825.81	601.49

Table 2: Eigen Value Matrix

	1	2	3	4	5	7
1	2380.540	-4.83E-13	-1.49E-13	-5.68E-13	7.96E-13	1.28E-13
2	-2.98E-13	381.5992	-2.71E-14	-3.55E-14	4.26E-14	-5.68E-14
3	-4.26E-14	5.33E-15	228.5227	-7.51E-14	1.51E-14	3.11E-15
4	-5.44E-13	-4.13E-14	-8.37E-14	48.02152	-3.91E-14	-4.34E-14
5	6.37E-13	8.97E-14	1.50E-14	-6.66E-14	18.99952	7.06E-14
6	6.58E-14	-4.56E-14	-1.38E-14	-3.44E-14	8.37E-14	3.634583

Table 3: Eigen Vector Matrix (PCA)

	1	2	3	4	5	7
1	0.365160	0.181682	0.333255	0.000253	0.696124	0.487852
2	0.115973	-0.449366	-0.745338	-0.337200	0.261144	0.217233
3	0.026275	-0.100793	-0.302580	0.937600	0.130732	0.037532
4	0.914628	0.017771	-0.047938	0.006305	-0.232504	-0.326716
5	0.123675	-0.042546	0.031013	0.056338	-0.612894	0.776607
6	-0.026016	0.867626	-0.488463	-0.063141	-0.020621	0.059488

Two methods were applied, one is a supervised method (MD) and one is unsupervised method (KM) clustering algorithm. Each method applied two times one on the original images, and the second on the PCA image. Classification time is very important factor in such processing; one of the advantages of

feature extraction is to reduce this time through reducing the variance in the original image.

V. CONCLUSION

The classification process showed that the classification result from manipulating PCA image is better in both terms of accuracy and the speed of classification process than the classification made on the original images. In terms of Supervised classification, research noticed that the supervised classification gives more accuracy than the unsupervised. And that comes from the fact that in supervised classification, which implies kind of experience, has a more options of manipulated values. As example in Supervised the user can determines the vectors and the number of classes

REFERENCES

- [1] Schowengerdt, Robert A. (2007). *Remote Sensing: Models and Methods for Image Processing* (Third ed.): Academic Press.
- [2] Richards, John A. (2013). *Remote Sensing Digital Image Analysis: An Introduction* (Fifth Ed.): Spring.
- [3] Zhang, Mingyang, Wang, Kelin, Zhang, Chunhua, Chen, Hongsong, Liu, Huiyu, Yue, Yuemin, . . . Qi, Xiangkun. (2011). Using the radial basis function network model to assess rocky desertification in northwest Guangxi, China. *Environmental Earth Sciences*, 62(1), 69-76.
- [4] Awad, Mohamad, Chehdi, Kacem, & Nasri, Ahmad. (2007). Multicomponent image segmentation using a genetic algorithm and artificial neural network. *Geoscience and Remote Sensing Letters, IEEE*, 4(4), 571-575.
- [5] Saha, Sriparna, & Bandyopadhyay, Sanghamitra. (2008). *Unsupervised pixel classification in satellite imagery using a new multiobjective symmetry based clustering approach*. Paper presented at the TENCON 2008-2008 IEEE Region 10 Conference.
- [6] Katiyar, SK. (2013). An N-dimensional approach towards object based classification of remotely sensed imagery. *arXiv preprint arXiv:1303.6619*.
- [7] Achary, Tinku, & Ray, Ajoy K. (2005). *Image Processing Principles and Applications*: Wiley-Interscience.
- [8] Oommen, Thomas, Misra, Debasmita, Twarakavi, Navin KC, Prakash, Anupma, Sahoo, Bhaskar, & Bandopadhyay, Sukumar. (2008). An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences*, 40(4), 409-424.
- [9] Otakei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12(1), 27-31.
- [10] Alrawi, K. R., Gonzales, C., & Arquero, A. (1999). Supervised ART-II: A new Neural Network Architecture, with Quicker Learning Algorithm, For Learning and Classification Multi-Valued Input Pattern. Paper presented at the 7th European Symposium on Artificial Neural Networks, Bruges, Belgium.
- [11] Rao, V. B., & Rao, H. V. (2003). *C++ Neural Network and Fuzzy Logic* (Second ed.): BPB Publication.
- [12] Specht, D. F. (1990). Probabilistic Neural Network and Polynomial Adline as Complementary Techniques for Classification. *IEEE Transaction on Neural Network*, 1(1), 111-121.
- [13] Yang, Xiao-Hua, Wang, Fu-Min, Huang, Jing-Feng, Wang, Jian-Wen, Wang, Ren-Chao, Shen, Zhang-Quan, & Wang, Xiu-Zhen. (2009). Comparison between radial basis function neural network and regression model for estimation of rice biophysical parameters using remote sensing. *Pedosphere*, 19(2), 176-188.
- [14] Patil, Jayamala K, & Kumar, Raj. (2011). Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research*, 2(2), 135-141.
- [15] Camps-Valls G, Tuia D, Bruzzone L, and Benediktsson J. (2014). Advances in Hyperspectral Image Classification. *IEEE Signal Processing Magazine* (January 2014).
- [16] Banerjee B., Mohan B. K., Chauduri S. & Mohanty J. (2014). A Novel Graph Matching Based Approach For Land-Cover Classification Of Multi-Temporal Images. Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing.
- [17] Roy M., Ghosh S. & Ghosh A. (2014). A Novel Approach for Change Detection of Remotely Sensed Images Using Semi-Supervised Multiple Classifier System. *Information Sciences*, volume 269, june 2014, pages 35-47, Elsevier.
- [18] Tewkesbury A. P., et al. (2015). A Critical Synthesis of Remotely Sensed Optical Image Change Detection Techniques. Elsevier, *Remote Sensing of Environment*.
- [19] Harikumar A., Kumar A., Stein A., Raju P. L. N. & Murthy Y. V. N. K. (2015). An Effective Hybrid Approach To Remote-Sensing Image Classification. *International Journal of Remote Sensing Volume 36, Issue 11, 2015*.
- [20] Yichung X., Ambing Z. & William W. (2015). Mapping Wetlands and Phragmites Using Publically Available Remotely Sensed Images. *Photogrammetric Engineering & Remote Sensing*, Number 1 / January 2015. <http://dx.doi.org/10.14358/PERS.81.1.69>
- [21] Al-Barhawe, Deya Hazem. (2007). *Vectorization of Remote Sensing Data for GIS Use*. (Ph.D), University of Mosul.
- [22] Al-Mihimdi, M. G. (2006). *Remote Sensing and Geographical Information Systems Techniques for Landuse Classification and Spatial Analysis for Adamiy* (Ph.D) , Baghdad-Iraq.
- [23] Aggarwal, Shefali. (2004). *Principle of Remote Sensing*. Paper presented at the Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, Dehra Dun, India.
- [24] Abrams, Michael, Curran, Paul, Dekker, Arnold, Dejong, Steven M., & Schaepman, Michael. (2005). *Remote Sensing Image Analysis: Including the Spatial Domain* (Vol. 5): Spring.
- [25] Cihlar, J., xiao, Qinghan, Chen, J., Beaubien, J., Fung, K., & Latifovic, R. (1998). Classification by progressive generalization: a new automated methodology for remote sensing multichannel data. *International Journal of Remote Sensing*, 19, 2685-2704.
- [26] David, J. F., Yan, K. Y., Elmagarmid, A. K., & Aref, W. D. (2001). Automatic image segmentation by integrating Color-Edge extraction and Seeded Region Growing. *IEEE Transaction on Image Processing*, 10(10), 1454-1466.