IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Performance Evaluation of MOA v/s KNN Classification Schemes: Case Study of Major Cities in the World

Sanjay Srivas[1*], P. G. Khot[2]

[1]Dept. of Electronics & Computer Science, RTM Nagpur University Nagpur, Maharashtra, India
[2]Dept. of Statistics, RTM Nagpur University Nagpur, Maharashtra, India

**Corresponding Author: sanjaysrivas.iit@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i4.489495 | Available online at: www.ijcseonline.org

Accepted: 15/Apr/2019, Published: 30/Apr/2019

***Abstract* —** Satellite imageries are widely available from various sources which can be used for Land use/Land cover analysis. Land use/Land cover analysis is necessary for environmental monitoring, urban planning and natural resource analysis. In this paper, we have used newly created algorithm- Multi Objective Algorithm (MOA) which is the combination of two metaheuristic algorithms for classification of satellite imageries. Classification result was compared with the KNN (K-Nearest Neighbour) algorithm. In this view, satellite imageries of Delhi and Shenyang were used for the experiment purpose. Also accuracy of classification was measured using the error matrix/kappa coefficient and was compared with the KNN classification technique. The classification results of the two major cities indicate a substantial difference in the percentage of overall accuracy and kappa coefficient value.

***Keywords* —** Classification, Land use/Land Cover, K-Nearest Neighbour, MOA, Accuracy Assessment.

I. INTRODUCTION

Remote sensing research focused on feature selection and has attracted the attention of the remote sensing community because feature selection is a prerequisite for image processing and various applications. Many feature selection methods have been proposed to improve the classification accuracy. Classification results of remotely sensed data are usually summarized as confusion matrices (contingency tables) [1]. However, the contingency tables are unable to assess classification accuracies completely because the tables do not provide the accuracies for each classification category. Therefore, we are in need of more performance metrics like producer's accuracy, user accuracy and Kappa Statistics. Various new classifiers have been developed as advanced techniques and are applied to the remote sensing field. Therefore, multiple comparisons of results from the new classifiers with those from the conventional classifiers are needed. The objective of this study is to compare the performance of the conventional KNN techniques with the proposed MOA algorithm [2], which was developed in the MATLAB tool.

II. RELATED WORK

Classification

Classification between the objects is an easy task for humans, but it is a complex problem for machines. Classification techniques are used in identifying to which of a set of categories (subpopulations) the new observation belongs. The data set used may be a bi-class (binary classification) or a multiclass and discrete in nature. In the language of the machine learning, it is the instance of supervised learning whereas in the case of the unsupervised learning it is called clustering. Classification technique is widely used in many sectors, but here we have limited our focus in the field of remote sensing related studies .In the remote sensing field, Satellite image classification is all about the grouping of pixels into meaningful classes depending upon its pixel values [3]. It involves the interpretation of remote sensing images, spatial data mining and studies various vegetation, urban developments and other land features to determine various land uses in an area. An algorithm that is used for the classification is called classifier. Sometimes the mathematical function of the algorithm is also referred to as the classifier. For extracting information and interpreting it from the digital satellite image, two widely known methodologies, Supervised Classification and Clustering (Unsupervised Classification) are adopted [4].

Supervised Classification

Supervised classification is the methodology mostly used for the quantitative analysis of remote sensing image data. Training of the sample data is most important in the study of the supervised classification. In supervised classification technique, the location of land cover types should be known prior. The supervised classification includes functionality such as analysing input data, creating training samples and signature files. Supervised classification is the method of known informational classes (training sets) to classify pixel of the unknown identity. Accuracy of the methods highly depends on the samples taken for training. Training samples are of two types, one used for classification and another for supervising classification accuracy [5].

Unsupervised Classification
Clustering analysis is also called as a segmentation analysis or taxonomy analysis. Clusters created by using these classification techniques are such that objects in the same cluster are having very similar characteristics and features whereas the objects in the other cluster are very distinct and dissimilar and that too without training the model. It is used to draw inferences from the datasets that have no labelled responses. They are used for the exploratory data analysis to find out the hidden patterns in the data [6].

Popular Classification Techniques:
Following are the most popular classification techniques which are generally used in Landsat images classification.

1. KNN (Nearest Neighbour)
2. Maximum Likelihood
3. K –Means

KNN (Nearest Neighbour) Classification
K-Nearest Neighbours (KNN) is one of the simplest machine learning algorithms, which is based on feature similarity. Selecting the correct value of 'k' is a process called parameter tuning and it is important for better accuracy. KNN is also referred to as a non-parametric machine learning algorithm. It stores all the available case and classifies new cases based on a similarity measure. This algorithm performs much better when the data size is large and linear in nature. To classify the feature vector within the feature space, it calculates the unidentified distance between the points. Commonly, Euclidean distance is used to calculate the distance [7].

Maximum Likelihood
The most common classification of the satellite images takes place with the help of extensive classification algorithm such as a maximum likelihood algorithm. It comes under the supervised classification methodology. The classifier guesses the probability with which a specific pixel belongs to a specific class. Maximum Likelihood Classification is based on the Bayes Classifications. Mean vector and covariance are the key component of the MLC that can be retrieved from training data. To summarize, maximum likelihood estimation is a method to seek the probability distribution that makes the observed data most likely [8].

K-Means
K-Means is the popular unsupervised algorithm originally used in the signal processing. K means method uses a vector quantization. The k parameter stands for the desired number of the cluster to generate. Each pixel in the image is assigned to the nearest cluster centre, which is recomputed as the centroid of all pixels related to that cluster. This process is repeated unless and until a desire of stopping criteria is reached. It is often used as a pre-processing step for other algorithm and it can be easily applied to even a large data set without any difficulties or complexities [9].

Performance Measure Techniques
The metrics that one chooses to evaluate one machine learning based classification model are very important. Choice of the metrics influences how the performance of the model is measured.

Confusion/Contingency/Error Matrix
This is the most commonly used accuracy assessment technique in classification. It is also known as contingency matrix or error matrix (Scofield et al., 2015). Overall accuracy, producer accuracy and user accuracy of classified image can be derived from the error matrix [10].

Producer Accuracy
Producer's accuracy measures errors of omission, which gives the prospective that how many of the pixels are correctly classified in that particular class.

User Accuracy
User's accuracy measures errors of commission, which provides the probability of a pixel classified on the map, which represents the actual class on the ground.

Overall Accuracy
Overall Accuracy is measured by dividing the correctly classified pixels by the total number pixel checked. Besides the overall accuracy, this technique can be used to determine the accuracy of the individual classes as the overall accuracies do not indicate how the accuracies are distributed across the individual classes [11].

Kappa Statistics
Kappa Cohen's k is the most common and extensively used measure of inter-rater reliability when the outcome of interest is measured on a nominal scale. This method is commonly used to measure the accuracy of image classification as well.

$$K = \frac{P_o - P_e}{1 - P_e}$$

Observed proportional agreement:

$$P_0 = \sum \frac{M_k}{N} = \frac{a+b}{n}$$

Expected proportional agreement:

$$P_e = \sum \frac{R_k C_k}{N^2} = \frac{N_1 \times N_2}{N^2} + \frac{N_2 \times N_4}{N^2} \quad \text{………… Eqn 3}$$

Where, N is the total number of observations and M_k is the number of agreements in the diagonal of confusion matrix. R_k and C_k are the rows and columns for the kth categories.

Kappa Coefficient:

Kappa coefficient, better known as Cohen's Kappa, is used to compare performance in the machine learning. Statistically, it measures the agreement between categorical variable X and Y [12].

It is generally considered to be a better measure than simple percent agreement calculation since k takes into account the agreement occurring by chance. This technique is also used to assess the agreement between alternative methods of categorical assessment when new techniques are under study. Kappa values can be interpreted in the following ways [13,14]:

Table 1. Kappa values interpretation

S. No.	Values observed in the range	Interpretation
1	< 0.2	Poor Agreement
2	0.2 to 0.4	Fair Agreement
3	0.4 to 0.6	Moderate Agreement
4	0.6 to 0.8	Good Agreement
5	0.8 to 1.0	Very Good

III. Experimental Results

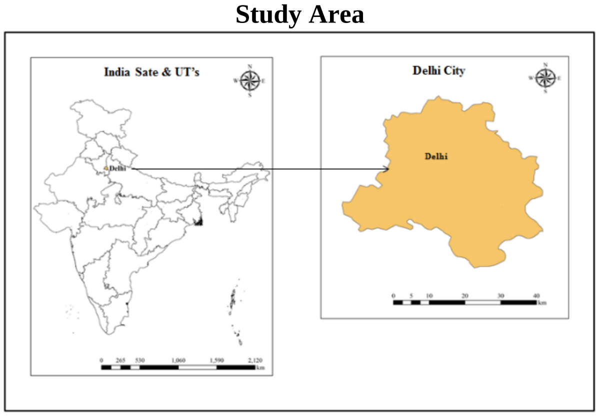
In this section, we performed a set of experiment on two major cities (Delhi and Shenyang) in the world. The experiments have been conducted using MOA Algorithm and the results are compared with the KNN classification method. Below is the name of the cities which were considered for experiment:

1. Delhi
2. Shenyang

A. Study Area (Delhi)

The study area, Delhi city, is the historical and political capital of India, located in the centre of India. The study area is approximately 5,333 sq.km. shown in Figure 1.

Study Area



Input Image

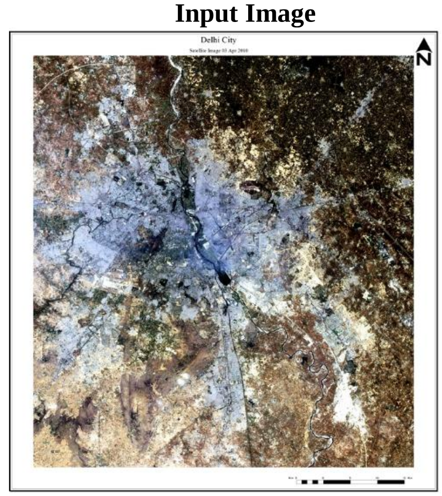


Figure 1: Highlighted Study Area & Input Image

Delhi City
03 Apr 2010 using KNN

Delhi City
03 Apr 2010 using MOA

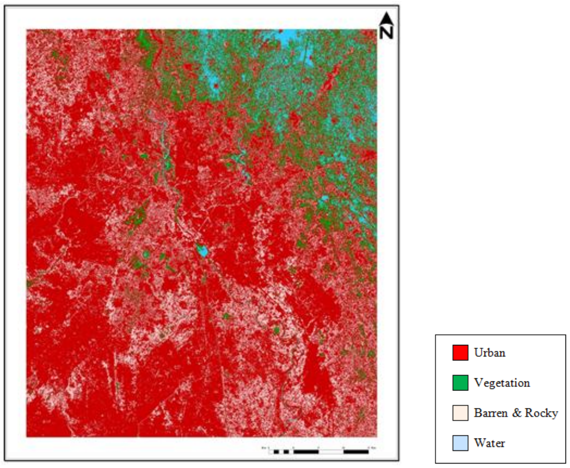


Figure 2(A): Classified Image of study area using KNN

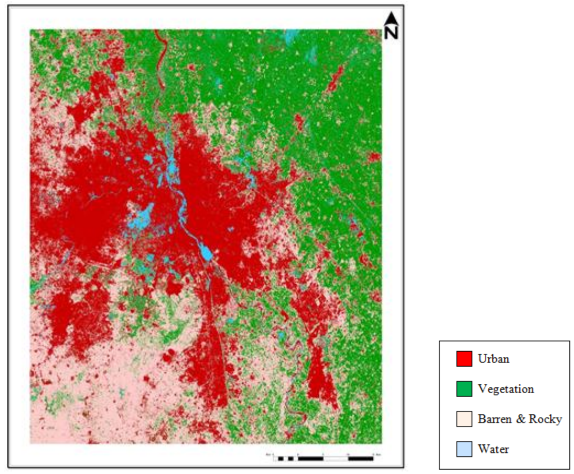


Figure 2(B): Classified Image of study area using MOA

For the Figure 2(A) classified image, there were a total 5,925,260 pixels. The area of each class was calculated based on the number of pixels classified into respective class. Thus, the number of pixels distribution and percentage are tabulated in Table 6.1. The percentage of pixels as classified are: Urban (66.15%), Vegetation (9.27%), Barren & Rocky (20.19%) and Water (4.40%).

For the Figure 2(B), a total of 5,333 sq.km. area was considered for image classification. The area of each class was calculated based on the number of pixels classified into respective class. The areas of classified image are: Urban (1,702.81 sq.km.), Vegetation (1,672.92 sq.km.), Barren & Rocky (1,796.65 sq.km.) and Water (160.48 sq.km.).

Error Matrix for Accuracy Assessment

Error matrix prepared to assess the classification accuracy. As per the Table 2, Error matrix prepared to assess the classification accuracy. As per the Table 2, classification using MOA was more accurate than KNN classifier. Overall accuracy by using KNN classifier and MOA Algorithm was 40% and 72 % respectively.

Table 2. Error matrix using KNN & MOA

	Classification using KNN						Classification using MOA					
Class	**A**	**B**	**C**	**D**	**Total**	**User Accuracy**	**A**	**B**	**C**	**D**	**Total**	**User Accuracy**
A	14	12	4	0	30	46.67%	18	1	2	1	22	81.82%
B	2	6	0	1	9	66.67%	1	6	2	1	10	60.00%
C	2	6	0	0	8	-	2	4	12	0	18	66.67%
D	1	2	0	0	3	-	0	0	0	0	0	-
Total	19	26	4	1	50		21	11	16	2	50	
Producer Accuracy	73.68%	23.08%	-	-			85.71%	54.55%	75.00%	-		

(A: Urban, B: Vegetation, C: Barren & Rocky D: Water)

B. Study Area (Shenyang)

The study area, Shenyang city, is the provincial capital and the largest city of Liaoning Province, People's Republic of China, as well as the largest city in Northeast China by urban population. The study area is approximately 3,198 sq.km. shown as below.

Input Image -06 Apr 2010

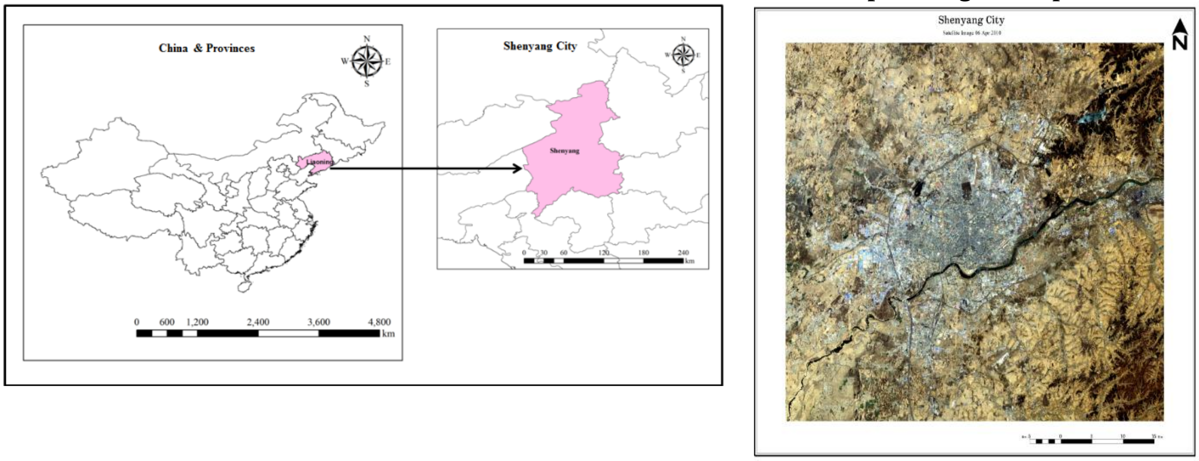


Figure 3: Highlighted Study Area & Input Image

Shenyang City
06 Apr 2010 using KNN

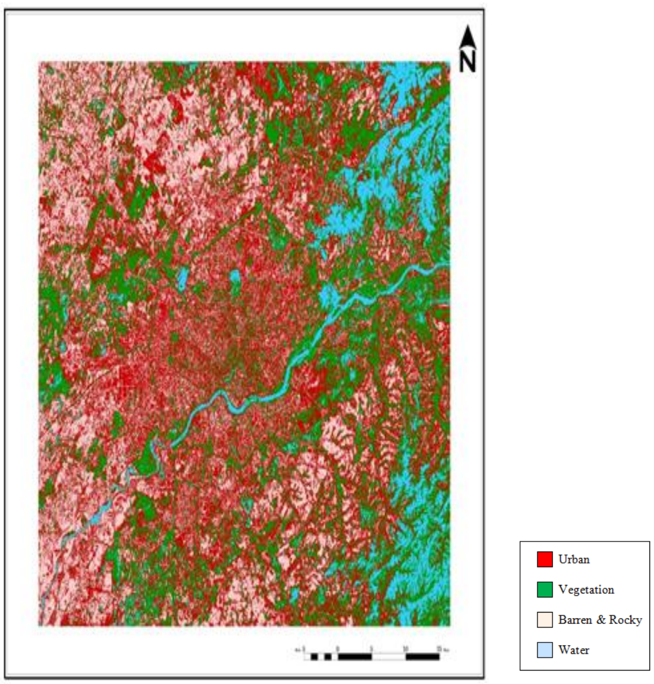


Figure 4 (A): Classified Image of study area using KNN

Shenyang City
06 Apr 2010 using MOA

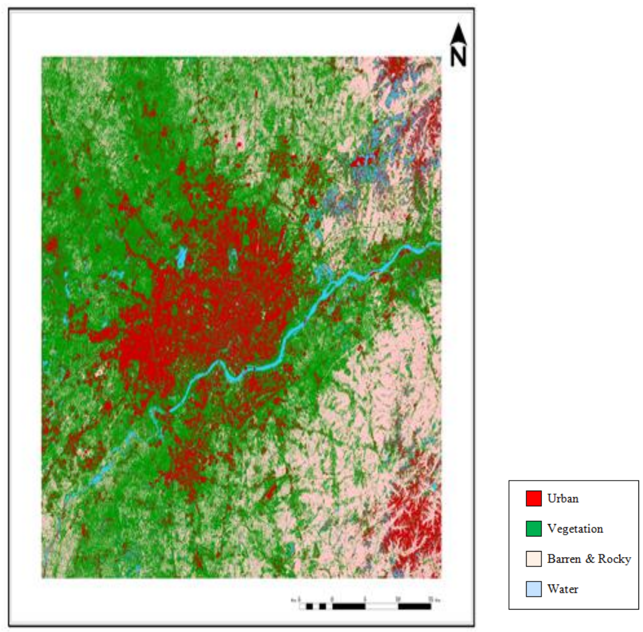


Figure 4 (B): Classified Image of study area using MOA

For the Figure 4(A) classified image, there were a total of 3,553,056 pixels. The area of each class was calculated based on the number of pixels classified into respective class. Thus, the number of pixels distribution and percentage are tabulated in Table 6.10. The percentage of pixels as classified are: Urban (39.67%), Vegetation (25.75%), Barren & Rocky (26.34%) and Water (8.24%).

For the Figure 4(B), a total of 3198 sq.km. area was considered for image classification. The area of each class was calculated based on the number of pixels classified into respective class. The areas of classified image are: Urban (1,562.01 sq.km.), Vegetation (625.54 sq.km.), Barren & Rocky (897.50 sq.km.) and Water (112.73 sq.km.).

Error Matrix for Accuracy Assessment

Error matrix prepared to assess the classification accuracy. As per the Table 3, classification using MOA was more accurate than KNN classifier. Overall accuracy by using KNN classifier and MOA Algorithm was 50% and 80% respectively.

Table 3. Error matrix using KNN & MOA

	Classification using KNN						Classification using MOA					
Class	**A**	**B**	**C**	**D**	**Total**	**User Accuracy**	**A**	**B**	**C**	**D**	**Total**	**User Accuracy**
A	13	0	10	0	23	56.52%	6	5	0	0	11	54.55%
B	0	0	0	0	0	-	0	18	3	0	21	85.71%
C	6	0	10	1	17	58.82%	0	2	15	0	17	88.24%
D	1	0	7	2	10	20.00%	0	0	0	1	1	100.00%
Total	20	0	27	3	50		6	25	18	1	50	
Producer Accuracy	65.00%	-	37.04%	66.67%			100.00%	72.00%	83.33%	100.00%		

(A: Urban, B: Vegetation, C: Barren & Rocky D: Water)

IV. COMPARATIVE ANALYSIS

The comparison of the overall accuracy values of each classified image using KNN classification and MOA algorithm are presented in Table 4. These values are calculated from Error Matrix Table 2 and Table 3.

Table 4. Accuracy assessment for the classified images

Study Area	Classification using KNN		Classification using MOA	
	Overall Classification Accuracy	Overall Kappa Statistics	Overall Classification Accuracy	Overall Kappa Statistics
Delhi	40.00 %	0.0969	72.00 %	0.5731
Shenyang	50.00 %	0.1940	80.00 %	0.6878

Furthermore, the comparison of both the classified images using KNN & MOA reveals that using MOA (optimised algorithm) outperforms as compared to KNN classification. It is also observed that overall classification accuracy is much higher using MOA algorithm.

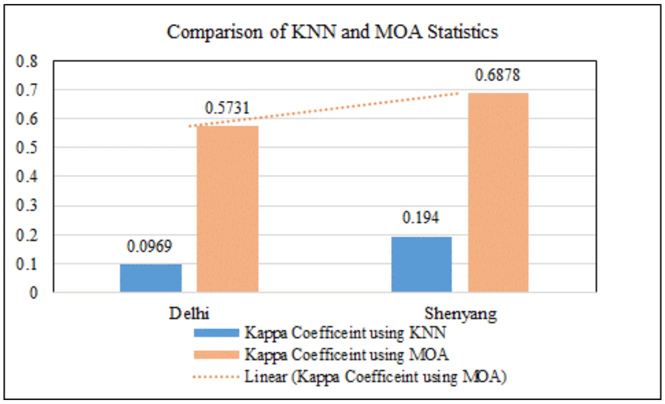


Figure 5: Graphical representation of Kappa Statistics

Figure 5 represents the comparison of Kappa Statistics obtained after performing the accuracy assessment after creating confusion matrix. It is clearly observed that the classification accuracy shown by MOA is very high in both the classified images, whereas the classification accuracy shown by KNN is very low. It can be inferred from this observation that MOA has a high level of accuracy as compared to KNN.

V. CONCLUSION

Accuracy assessment is one of the most important aspects of image classification. In this paper, we have performed experiments using Multi Objective Algorithm, which is an integration of two Meta heuristic algorithms (i.e. CSA and DA). Its idea is to reduce the computation while distributing the pixels into different classes as per the criteria. We have compared the MOA algorithm with the existing KNN classification method. The algorithm was analysed for a number of experimental problems and compared with KNN classification method. The results indicate that MOA is outperforming the KNN classification method while considering the accuracy assessment. We have achieved outperforming results from the new statistical method with a high accuracy rate.

REFERENCES

[1] Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46.

[2] Sanjay Srivas, P. G. Khot, (2018). Analysis & Visualization of Multidimensional GIS Images Using Multi Objective Algorithm (MOA). *International Journal of Computer Sciences and Engineering*, 6(8), 460-464.

[3] Darius Phiri, & Justin Morgenroth. (2017). Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sensing*, *9*(9), 967.

[4] Liu, X. *Supervised Classification and Unsupervised Classification*. 12.

[5] Kalra, K., Goswami, A. K., & Gupta, R. (2013). A Comparative Study of Supervised Image Classification Algorithms for Satellite Images. *International Journal of Electrical, Electronics and Data Communication*, 1(10), 7.

[6] Canty, M. J., & Nielsen, A. A. (2006). Visualization and unsupervised classification of changes in multispectral satellite imagery. *International Journal of Remote Sensing*, *27*(18), 3961–3975.

[7] Cai, Y., Ji, D., & Cai, D. (2010). A KNN Research Paper Classification Method Based on Shared Nearest Neighbor. *NTCIR*, 5.

[8] Ahmad, A., & Quegan, S. (2012). Analysis of Maximum Likelihood Classification on Multispectral Data. *Applied Mathematical Sciences*, 6(129), 6425 - 6436.

[9] Yadav, J., & Sharma, M. (2013). A Review of K-mean Algorithm. *International Journal of Engineering Trends and Technology*, 4(7), 5.

[10] Scofield, G. B., Pantaleao, E., & Negri, R. G. (2015). A Comparison of Accuracy Measures for Remote Sensing Image Classification: Case Study In An Amazonian Region Using Support Vector Machine, *International Journal Image Processing*, 9(1), 11–21.

[11] Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, 08(04), 611–622.

[12] Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, 37(5), 360-363.

[13] Olmsted, A. Calculating kappa measures of agreement and standard errors using SAS software: some tricks and traps. 6

[14] Bharatkar, P. S., & Patel, R. (2013). Approach to Accuracy Assessment tor RS Image Classification Techniques. *International Journal of Scientific & Engineering Research*, *4*(12), 8.

