Classification of Intermediate Sites using IRS-Panchromatic Satellite Imagery

Sinchana G S, Choodarathnakara A L and Neelappa

Government Engineering College, Kushalnagar, Karnataka, INDIA

ABSTRACT

Remotely Sensed data is an important component of Land Use/Land Cover (LU/LC) studies. This paper compares the performance of ISODATA classification with Mahalanobis Distance classification for Arsikere semi-urban study area of Karnataka State, INDIA using IRS-P5 PANF satellite imagery. The Arsikere study area is an intermediate region comprised primarily of water features mixed with impervious features. During the beginning of experiment, ISODATA unsupervised classification was applied on IRS data. Later, Mahalanobis Distance classification was applied on IRS data with 6000 training sites and 100 validation points for water, vegetation, soil and impervious surface features which were randomly generated using a stratified sampling approach. The LU/LC data associated with these points were then compared with Topographic Maps (Survey of India, No. D43Q3, D43Q7) and Ground Truth Data for performance analysis. Based on the confusion matrices obtained for the sample set, the OCA, Kappa statistics were compared with ISODATA. The experimental analysis shows that unsupervised ISODATA classification provides accuracy of 84% in Arsikere, semi-urban area, however Mahalanobis distance classification give up 92% OCA with TS = 6000 and VS = 100.

Keywords

Remote Sensing, LU/LC Features, ISODATA, Mahalanobis Distance Classification.

1. INTRODUCTION

Remote sensing is defined as the measurement of object properties on the Earth's surface using data acquired from aircraft and/or satellite platform. Such, remote-sensing data consist of discrete point measurements or a profile along a flight path, which are most interested in measurements over a two-dimensional spatial grid. Images of Remote-sensing systems particularly those deployed on satellites provide a repetitive and consistent view of the Earth that is invaluable for monitoring short-term and long-term changes as well as the impact of human activities. Some of the important applications of remote-sensing technology are: environmental assessment and monitoring (urban growth, hazardous waste), global change detection and monitoring (atmospheric ozone depletion, deforestation, global warming), agriculture (crop condition, yield prediction, soil erosion), nonrenewable resource exploration (minerals, oil, natural gas), renewable natural resources (wetlands, soils, forests, oceans), meteorology (atmosphere dynamics, weather prediction), mapping (topography, land use, civil engineering), news media (illustrations, analysis), military surveillance and reconnaissance (strategic policy, tactical assessment).

Thomas N. Lillesand [1] has explained the basic concepts and elements necessary to conceptualize an ideal remote sensing platforms and applications. Dr. B. C. Panda [2] has provided concepts of Remote Sensing, essential components of Remote Sensing and types of Remote Sensing. D. Lu and Q. Weng [3] elucidate image classification process and advanced image classification techniques for improving classification accuracy.

Supervised classification classifies pixels based on known properties of each cover type and it requires representative land cover information in the form of training pixels. The minimum distance classifier employs the central values of the spectral data that forms the training data set to classify pixels. The neural network classification is a self-adaptive method that is able to estimate the posterior probabilities, which provide a basis for establishing the classification rule [4], [5]. The support vector machine method involves a learning process based on structural risk minimization, which can minimize classification error without the need to assume data distribution [6].

Dongshui Zhang et al. [7] has designed and implemented a classifier named Maximum Gray Slope Correlation classification based on the gray slope correlation degree model and the remote sensing classification mechanism. Authors have conducted the comparative classification tests between the gray relational classification and other conventional remote sensing classification methods using small samples. A. L. Choodarathnakara et al. [12] proposed PCA method to detect built-up features using LANDSAT 7 ETM+ Satellite Imagery. In this method, first three components PCA1, PCA2 and PCA3 were fused to get PCA1+PCA2+PCA3 with 98% of six dimensions B1, B2, B3, B4, B5, B7 and PCA model was successful with 98% accuracy. Ashok Kumar T [8] presented the performance and employability of the decision tree classification algorithm in respect of varying training dataset size for class hierarchy levels I and II along with effects of ancillary data on tree complexity with number of rules induced.

A. L. Choodarathnakara [9] was taken up research work with the objective of designing an efficient and reliable classification strategy in an attempt to find answers to some of the conflicting issues dealt within the existing literature pertaining to classification of fine resolution RS data. The authors have conducted experiment on the MS data of IRS LISS-IV sensor of 5m spatial resolution and PAN data of 2.5m spatial resolution. Authors have concluded that the hard classification procedure fails to classify mixed pixel problem in Arasikere semi-urban area of Karnataka State, INDIA. To overcome this problem, authors proposed Decision Tree technique along with Mamdani_Fuzzy Inference System (M_FIS) as a hybrid classifier and concluded that Mamdani_FIS was a powerful candidate to classify mixed pixels present in semi-urban areas.

Tushar Rajendra Baviskar and B. D. Jadhav [10] carried out a study on change detection of the Pune City over the period of 1999-2015. Accuracy assessment detects the changes on Earth surface and the results of proposed method shows that settlement of the study area has been increased and agriculture land, vegetation, fallow land have been decreased. In this study different classification algorithms have been studied such as K-mean, ISODATA, Parallelpiped, MLC, Spectral Angle Mapper. Authors conclude that SAM gives most accurate result after validating accuracy of all classifiers. Giles M. Foody [11] has explained the background and methods of classification accuracy assessment that are commonly used and recommended. Foody elucidated different types of errors encountered in an image classification and concluded that the value of thematic map is a function of accuracy of the classification and the assessment of classification accuracy is not a simple task.

The objective of this research work is to assess and compare the accuracy of unsupervised ISODATA classification with supervised Mahalanobis Distance classification using IRS Satellite Imagery. This investigation analyzed the semi-urban study area with intermediate land use/land cover compositions for urban planning purpose. The rest of the paper is organized as follows, Section I contain the Introduction about Remote Sensing and its applications, Section II contain RS Data Classification, Section III contain the study area and the methodology proposed, Section IV contain Result Analysis, Section V conclude the final comments on the research work.

2. RS DATA CLASSIFICATION

Remote sensing is the science of acquiring information about the Earth's surface without actually being in contact with it. Here the device is a remote sensing sensor that is operated from air-borne and space-borne platforms to assist in inventorying, mapping and monitoring earth resources. Remote Sensing Image Classification is a process of automatically categorizing all pixels in an image into finite number of classes or themes. Multi-spectral classification is an information extraction process that analyses these spectral signatures and assign the pixels to the classes based on similar signatures.

2.1 Supervised Image Classification

In supervised image classification, the image analyst "supervises" pixel categorization process by specifying the computer algorithm to numerical descriptors of the various land cover types present in a scene. To do this, representative sample sites of known cover type called training areas are used to compile a numerical "interpretation key" that describes the spectral attributes for feature type of interest. Each pixel in the data set is then compared numerically to each category in the interpretation key and labeled. There are a number of numerical strategies that can be employed to make this comparison between unknown pixels and training set pixels namely, Maximum likelihood, Minimum Distance, Mahalonobis Distance, etc., [13, 14].

2.1.1 Mahalanobis Distance:

The Mahalanobis distance algorithm assumes that the histograms of the bands have normal distributions. In Mahalanobis Distance classification the class signature will be in the form of class mean vectors and the covariance matrices so that clusters that are highly varied lead to similarly varied classes and vice versa. However, the disadvantage is that the derived classes may not be statistically separable. The Mahalanobis distance uses statistics for each class but assumes that all class covariance are equal. All pixels are classified to the closest region of interest (ROI) class, depending on the distance threshold specified by users; some pixels may be unclassified if they do not meet the threshold [4].

2.2 Unsupervised Image Classification

Unsupervised image classification (commonly referred to as clustering) is an effective method of partitioning remote sensor image data in multispectral feature space and extracting land-cover information. Compared to supervised classification, unsupervised classification normally requires only a minimal amount of initial input from the analyst. This is because clustering does not normally require training data. The unsupervised procedures are applied in two separate steps. In the first step the image data are classified by aggregating them into the natural spectral groupings or clusters. Then the image analyst determines the land-cover identity of these spectral groups by comparing the classified image data to ground reference data.

2.2.1 ISODATA Classification:

ISODATA is iterative in that it repeatedly performs an entire classification (outputting a thematic raster layer) and recalculates statistics. Self-Organizing refers to the way in which it locates clusters with minimum user input. The ISODATA method uses minimum spectral distance to assign a cluster for each candidate pixel. The process begins with a specified number of arbitrary cluster means or the means of existing signatures and then it processes repetitively so that those means shift to the means of the clusters in the data. Since, the ISODATA method is iterative; it is not biased to the top of the data file, as the one-pass clustering algorithm.

2.3 Accuracy Assessment

No classification is complete until its accuracy has been assessed. In this context, the "accuracy" means the level of agreement between labels assigned by the classifier and the class allocations on the ground collected by the user. There are many kinds of accuracy assessment techniques like spatial accuracy, thematic accuracy, temporal accuracy and topological accuracy. The following are the most commonly used methods to do the accuracy assessment.

2.3.1 Overall Classification Accuracy:

Overall accuracy is the proportion of all reference pixels, which are classified correctly. It is computed by dividing the total number of correctly classified pixels (the sum of elements along the main diagonal) by the total number of reference pixels. According to the error matrix, the overall accuracy can be calculated as:

$$\text{OCA}=\frac{\sum_{K=1}^{N} a_{kk}}{\sum_{i,k=1}^{N} a_{ik}}=\frac{1}{n}\sum_{K=1}^{N} a_{kk} \qquad (1)$$

2.3.2 Producer's Accuracy

Producer's accuracy tells how well the classification agrees with reference classification. The producer's accuracy can be calculated as:

$$\text{PA (class I)} = \frac{a_{ii}}{\sum_{i=1}^{N} a_{ki}} \qquad (2)$$

2.3.3 User's Accuracy

User's accuracy predicts the probability that a pixel classified as class I is actually belonging to class I. The user's accuracy can be calculated as:

$$\text{UA (class I)} = \frac{a_{ii}}{\sum_{i=1}^{N} a_{ik}} \qquad (3)$$

2.3.4 Kappa Statistics

Kappa Statistic is based on the difference between the actual agreement in the error matrix and the chance agreement. It is the agreement between the remotely sensed classification and the reference data as indicated by the major diagonal in the confusion matrix.

3. STUDY AREA AND METHODOLOGY

3.1 Satellite Data Products and Study Area

The data product used in this study is the Panchromatic RS image of IRS-P5 Cartosat-I satellite which has been launched and further supervised by ISRO. This satellite data product was procured from the NRSC, Hyderabad, India. Table I provides the specifications of satellite data utilized for the purpose of semi-urban study. The study area considered for this research work is semi-urban area of Arsikere taluk, situated in Hassan district of Karnataka State, India with geographical coordinates of 13° 18' 50" North, 76° 15' 22" East and with original name ARASIYA KERE. Fig. 1 shows the satellite image of Arsikere semi-urban study area of Hassan district, Karnataka State.

Table I. Details of the satellite data products used in this study

Satellite and Data Type	Date of Acquisition	Spectral Resolution	Spatial Resolution
IRS-P5 PANF	04/04/2011	0.55-0.85 µm	2.5 m

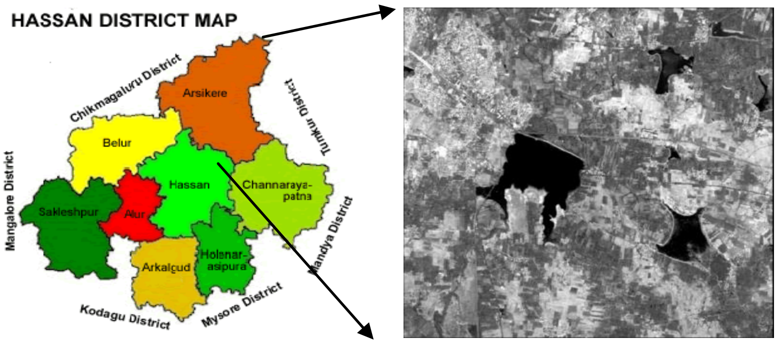


Fig 1: Arsikere semi-urban study area of Hassan District, Karnataka State, INDIA

3.2 Proposed Methodology

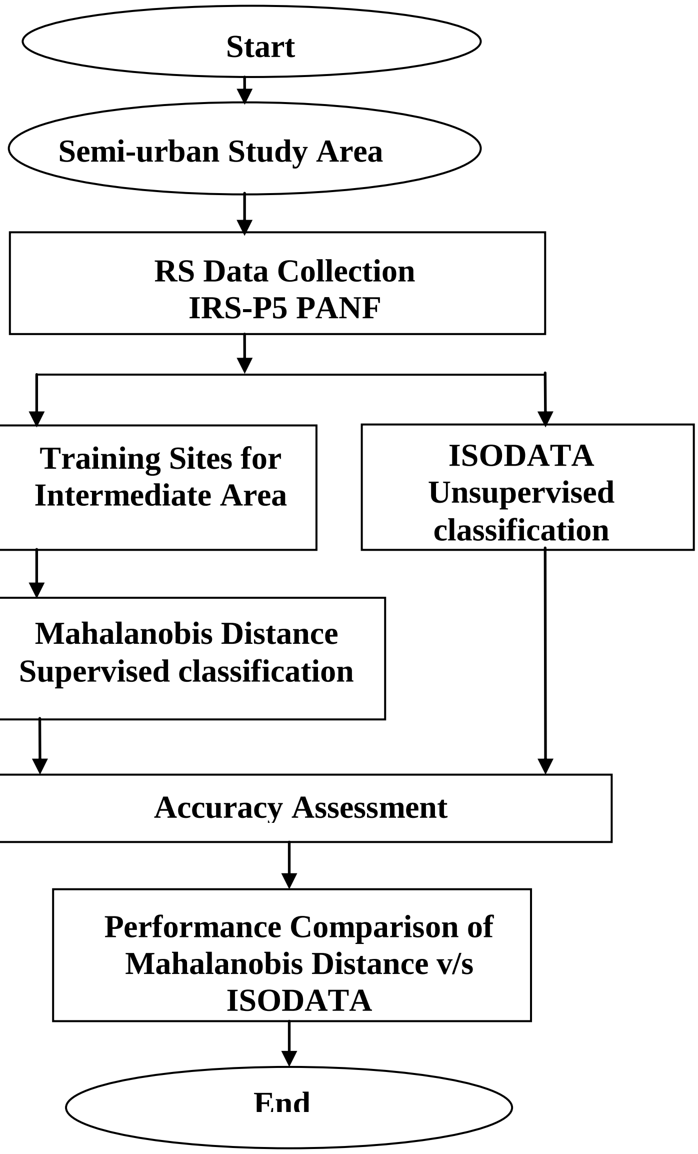


Fig 2: Proposed Methodology for classification of LU/LC features in Arsikere Semi-urban

The township of Arsikere study area is undergoing lot of changes being a semi-urban area between city and village land. This place is connected to various important cities in the state via bus and rail transport. In order to plan such a semi-urban land, the accurate classification of land use/ land cover features is necessary. In this context, the methodology is proposed which has been depicted in Fig. 2. During the first phase of the work the RS data was procured from NRSC Hyderabad. In the second phase, intermediate sites are identified for generating training site samples of water, vegetation, soil and impervious features of about 6000 TS. By employing Mahalanobis distance supervised classification confusion matrix was analyzed for four classes with 100 validation points for intermediate sites. The performance comparison of ISODATA versus Mahalanobis distance classification was dealt with the help of OCA and cross table as well.

4. EXPERIMENTAL RESULT ANALYSIS

The classification of geo-coded 2.5m spatial resolution data has been made using unsupervised ISODATA and supervised Mahalanobis Distance algorithms. Evaluation of these classifiers is dealt with confusion matrix based on accuracy assessment carried out using ERDAS IMAGINE V 9.2 RS image processing software.

4.1 Mahalanobis Distance Supervised Classification

Fig 3 shows the Arsikere semi-urban study area with intermediate training sites for experimentation. Fig 4 depicts the supervised Mahalanobis Distance Classified Image for intermediate training samples.

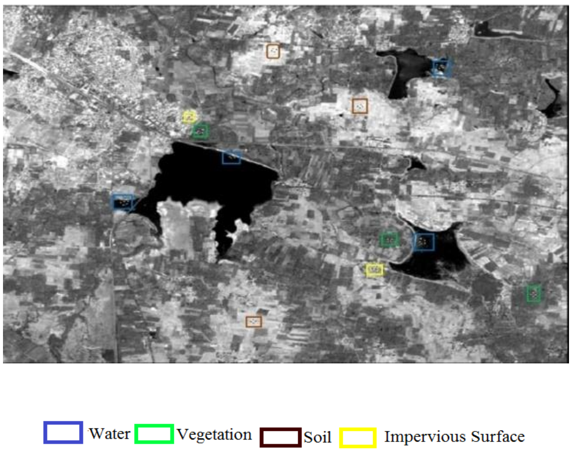


Fig 3: Intermediate Training Sites of the Arasikere semi-urban study area

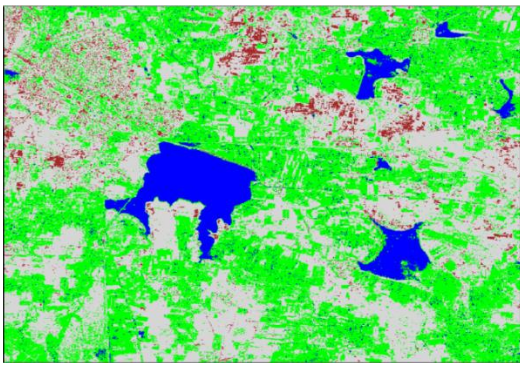


Fig 4: Supervised Mahalanobis Distance Classified Image

Table II. Confusion Matrix & Kappa Values obtained for Mahalanobis Distance Supervised Classification for Four Classes with TS = 6000 and VS = 100

Classes	1	2	3	4	Row Total	UA%
1	5	2			7	71.43
2		38		4	42	90.48
3			4		4	100
4		2		45	47	95.74
Column Total	5	42	4	49	92	
PA%	100	90.48	100	91.84		OCA: 92.00%
Kappa	0.6992	0.8358	1.0000	0.9166		OKS: 0.8640

Class Legends: 1: Water; 2: Vegetation; 3: Soil; 4: Impervious surface

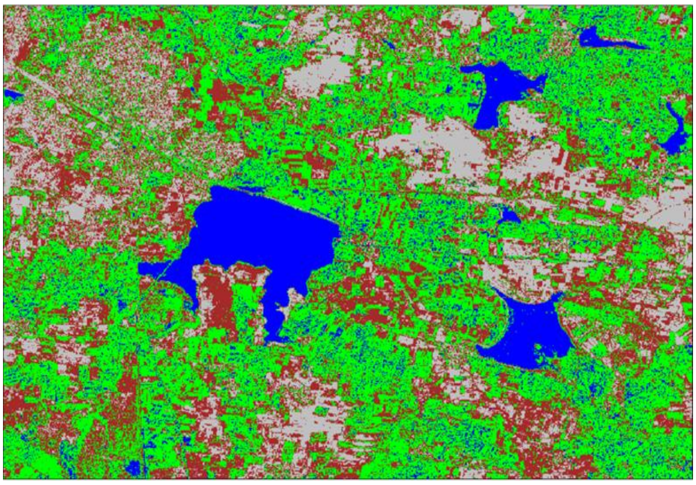
In Table II, out of 42 reference pixels of vegetation, 38 are correctly classified as vegetation and the rest 9.52% are misclassified to water and impervious surface producing a PA of 90.48% for vegetation. In other words out of the total 42 pixels which are classified as vegetation on the image, only 38 pixels represent vegetation and produce an UA of 90.48%. The remaining 9.52% of the pixels which are classified as vegetation are the misclassified pixels from other classes.

Out of 49 reference pixels of impervious surface 45 are correctly classified as impervious surface and the rest 8.16% are misclassified to vegetation producing a PA of 91.84%. In other words out of the total 47 pixels which are classified as impervious surface on the image, only 45 pixels represent impervious surface and produce an UA of 95.74%. The remaining 4.26% of the pixels which are classified as vegetation are the misclassified pixels from other classes.

Overall, Table II shows all the five pixels of water are classified as water by producing PA of 100% but remaining 28.57% of the pixels which are classified as water are misclassified from vegetation class. Also, all the four pixels of soil are correctly classified as soil by producing a PA and UA of 100%.

4.2 ISODATA Unsupervised Classification

Fig. 5 shows the Unsupervised ISODATA Classified Image for the Arsikere semi-urban area. The Table III depicts, Confusion Matrix & Kappa Values obtained for ISODATA unsupervised classification for four classes with 100 validation points.



Classes	1	2	3	4	Row Total	UA%
	10					

Fig 5:

ISODATA Classified Image for Intermediate Study Area

Table III. Confusion Matrix & Kappa Values obtained for ISODATA Unsupervised Classification for four Classes with 100 Validation Points

1		1			11	90.91
	5					
2		**35**	1	1	42	83.33
3		3	**24**	2	29	82.76
4		1	**2**	**15**	18	83.33
Column Total	15	40	27	18	**84**	
PA%	66.67	87.50	88.89	83.33		**OCA: 84.00%**
Kappa	0.893	0.722	0.763	0.796		**OKS: 0.773**

Class Legends: 1: Water; 2: Vegetation; 3: Soil; 4: Impervious surface

In Table III, out of 15 reference pixels of water, ten are correctly classified as water and the rest 33.33% are misclassified to vegetation class producing a PA of 66.67% for water. In other words, out of the total 11 pixels which are correctly classified as water on the image, only ten pixels represents water and produce an UA of 90.91%. The remaining 9.09% of the pixels which are classified as water are the misclassified pixels from vegetation class.

For vegetation feature, out of 40 reference pixels of vegetation, 35 are correctly classified as vegetation and the rest 12.5% are misclassified to water, soil and impervious surface classes producing a PA of 87.50%. In other words, out of the total 42 pixels which are correctly classified as vegetation on the image, only 35 pixels represents water and produce an UA of 83.33%. The remaining 16.68% of the pixels which are classified as water are the misclassified pixels from water, soil and impervious surface classes.

For soil feature, out of 27 reference pixels of soil, 24 are correctly classified as soil and the rest 11.11% are misclassified to vegetation and impervious surface classes producing a PA of 88.89%. In other words, out of the total 29 pixels which are correctly classified as soil on the image, only 24 pixels represents soil and produce an UA of 82.76%. The remaining 17.24% of the pixels which are classified as soil are the misclassified pixels from vegetation and impervious surface classes.

For impervious surface, out of 18 reference pixels of impervious surface, 15 are correctly classified as impervious surface and the rest 16.77% are misclassified to vegetation and soil classes producing a PA of 83.33%. In other words, out of the total 18 pixels which are correctly classified as impervious surface on the image, only 15 pixels represents impervious surface and produce an UA of 83.33%. The remaining 16.67% of the pixels which are classified as impervious surface are the misclassified pixels from vegetation and soil classes.

4.3 Comparison of ISODATA versus Mahalanobis Distance

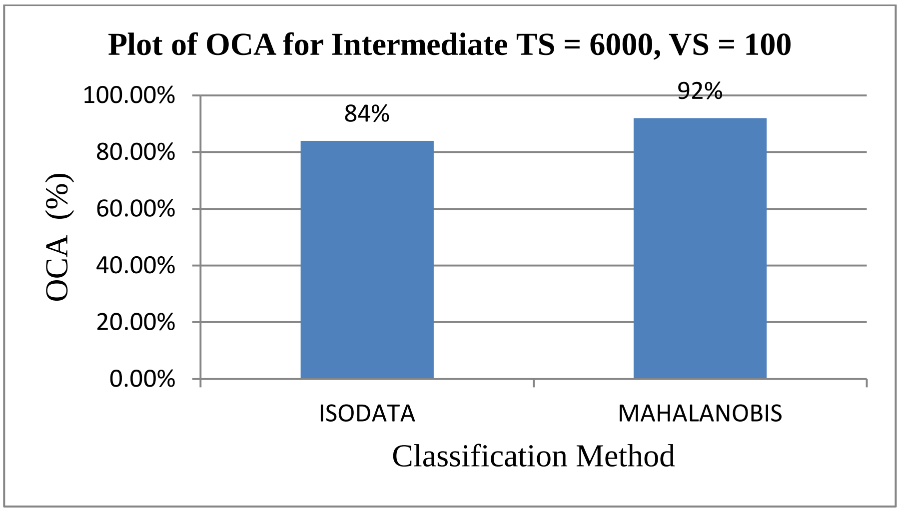


Fig 6: Plot of OCA for Intermediate Training Sites for TS = 6000 with VS = 100

Table IV. Cross Table between ISODATA and Mahalanobis Distance for Intermediate TS = 6000 with VS = 100

ISODATA	Mahalanobis Distance			
	Water	***Vegetation***	***Soil***	***Impervious surface***
Water	**0.5**			
Vegetation		**1.0857**		
Soil			**0.1666**	
Impervious surface				**3**

Fig 6 presents comparative graph of OCA for intermediate training sites for TS = 6000 with VS = 100 using Supervised Mahalanobis Distance classification and Unsupervised ISODATA classification. Table IV depicts the cross table of ISODATA versus Mahalanobis Distance classification. The comparative result demonstrates that the water class estimated by Mahalanobis Distance was 0.5 times the water estimated by ISODATA. The vegetation class estimated by Mahalanobis Distance was 1.08 times the vegetation estimated by ISODATA. The soil class estimated by Mahalanobis Distance was 0.16 times the soil estimated by ISODATA. The impervious surface class estimated by Mahalanobis Distance was 3 times the impervious surface estimated by ISODATA.

5. CONCLUSION

In this work, the study area considered is Arsikere taluk in Hassan district which was a semiconducting area with moderate rainfall. The experimental result concludes that ISODATA classification provides 84% OCA for Arasikere semi-urban area but Mahalanobis distance give in 92% OCA with TS = 6000 and VS = 100. Moreover, the satellite data used in this study consisting of only one band and hence it is not possible to classify more land use/land cover features. Therefore, the work can be continued by procuring high spatial and spectral resolution data with more number of bands. The study area is a semi-urban area consisting of mixed pixels and hence advanced classification approaches can be performed for better classification accuracy.

ACKNOWLEDGMENT

The authors would like to graciously thank National Remote Sensing Centre (NRSC), Hyderabad, INDIA for providing the data product for the study. Thanks to the Survey of India (SOI) for providing the Topographic Maps for this study and Karnataka State Remote Sensing Application Centre (KSRSAC), Bangalore. Acknowledgements due towards beloved students Kiran H T USN: 4GL14EC023, Sudhan A T USN: 4GL14EC055, Indresha C A USN: 4GL15EC402 and Shalini M C USN: 4GL15EC407. Special thanks to teaching and non-teaching faculty of GEC, Kushalnagar for their encouragement and guidance.

REFERENCES

[1] Thomas M Lillesand, Ralf W Kiefer and Jonathan W Chipman, 2004. "Remote Sensing and Image Interpretation", 5th Edition, John Wiley & Sons.

[2] Dr. B C Panda, 2005. "Remote Sensing Principles & Applications", Viva Publications.

[3] D.LU and Q.Weng, 2007. "A Survey of Image Classification Methods and Techniques for Improving Classification Performance", International Journal of Remote Sensing, Vol. 28, No.5, pp. 823-870.

[4] J.A. Richards, 1999. "Remote Sensing Digital Image Analysis: An Introduction", Springer Verlag: Berlin, Germany.

[5] G.P. Zhang, 2000. "Neural Networks for Classification: A Survey". IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, Vol. 30, No. 4, pp. 451-462.

[6] G. Mountrakis, J. Im and C. Ogole, 2011. "Support Vector Machines in Remote Sensing: A Review", ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 66, No. 3, pp. 247-259.

[7] Dongshui Zhang, Xinbao Chen , Y ongshun Han, Lixia Cong, Qinmin Wang and Xiaoqin Wang, 2014. "Remote Sensing Image Classification With Small Training Samples Based On Grey Theory", Third International Workshop on Earth Observation and Remote Sensing Applications, IEEE.

[8] Ashok Kumar T, 2010. "Advanced Image Processing Techniques and Algorithms For Classification of Semi-Urban Land Use/ Land Cover Features using High Resolution Satellite Data", Ph.D. Thesis, Visvesvaraya Technological University.

[9] A L Choodarathnakara, 2018. "A Soft Computing Approach For Classification Of Land Use / Land Cover Features In Semi-Urban Area Using High Resolution Satellite Imagery", Ph.D. Thesis, Visvesvaraya Technological University.

[10] Mr. Tushar Rajendra Baviskar and Mr. B.D.Jadhav, 2016. "Change Detection of Urban Area using Remote Sensing", IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT), pp. 28-32.

[11] Giles M. Foody, 2002. "Status of Land Cover Classification Accuracy Assessment", Remote Sensing of Environment, pp. 185-201.

[12] A L Choodarathnakara, Sinchana G S, Sowkya H K and Ramya B R, 2018. "Automated Detection of Built-up Feature in Bangalore City using LANDSAT 7 ETM+ Satellite Imagery", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 7, Issue 4, pp. 292-301.

[13] Barandela R and M Juraez, 2002. "Supervised Classification of Remotely Sensed Data with Ongoing Learning Capability", IJRS, Vol. 23, No. 22, pp. 4965-4970.

[14] Franklin S E and Wulder M A, 2002. "Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas", Progress in Physical Geography, 26, pp. 173–205.