ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

KNN and SVM based Satellite Image Classification

Sayali M. Jog[1], Mrudul Dixit[2], Anuja Rajgopalan[3], S. D. Ranade[4]

MKSSS's Cummins College of Engineering, Pune [1]

Assistant Professor, MKSSS's Cummins College of Engineering, Pune [2]

Scientist 'B', Central Water and Power Research Station, Khadakwasla, Pune[3]

Scientist 'D', Central Water and Power Research Station, Khadakwasla, Pune[4]

Abstract: Remote sensing is the art of acquiring information about an object or area using machine or device that is not physically connected to the object. Geology, urban planning, soil assessment and land cover/land use are the different applications of remote sensing. Remote sensing is widely used for generation of classification map. Image classification is used to group the pixels present in an image into different classes. This paper presents Support Vector Machine (SVM) and K Nearest Neighbor (KNN) based classification system for Indian Remote Sensing (IRS) satellite images. The proposed system consists of image enhancement, segmentation, selection of training data and classification. For image enhancement adaptive histogram equalization is used. Image segmentation is carried out using K means and Fuzzy C Means (FCM) clustering. Linear Multi-Class SVM (MCSVM) and KNN techniques are used for classification of remotely sensed imagery.

Keywords: Classification, Clustering, K Nearest Neighbour, Multi-Class Support Vector Machine, Satellite imagery, Remote sensing.

I. INTRODUCTION

Remote sensing employs the use electromagnetic energy in the form of heat, radio waves and light for measuring and detecting characteristics of target. Satellite and aircraft are the platforms used in remote sensing. In remote sensing energy measurement takes place in microwave region, visible light and emitted and reflected infrared regions of electromagnetic spectrum. Remote sensing involves use of hyperspectral or multispectral satellite images. Remotely sensed images are used in costal monitoring, geology, creation of classification map, ocean monitoring, agriculture, land cover/ land use, etc. [1].

The spatial and spectral resolution images provide the detail information about the target. These images are widely used in the field of habitat management, agriculture, land cover mapping and urban planning. The spatial details in such imagery are limited by number of pixels present in an image. The classification is used to group pixels present in an image into one of the land cover classes [2]. Classification technique is use to analyse the digital image and extract information from that image depending upon application [3].

The process of producing the thematic maps with themes like vegetation types, land use and geology is called image classification [3]. In remote sensing different quality images are produced depending upon type of sensor. Classification accuracy depends on image quality.

Image classification can be carried out using two ways: supervised classification and supervised classification.

In supervised classification it is necessary to have prior knowledge of study area. From training area features are extracted in supervised classification. Different supervised classifiers are: SVM, minimum distance classifier, K nearest neighbor classifier, maximum likelihood classifier, etc.

In unsupervised classification prior knowledge of study area is not needed. For selection of training data image analyst is not required. Unsupervised classifier is more automated and elegant as compared to supervised classifier. Different supervised classifiers are: k means clustering, fuzzy k means clustering, etc.

This paper contains following sections: Section II describes fuzzy C means and K means clustering. Section III introduces SVM. Section IV describes KNN classifier. Section V consists of result of classification. Paper concludes in Section VI.

II. CLUSTERING TECHNIQUES

Grouping of observations/ data into few segments is called clustering. Clustering algorithm segments data in such a way that data within segments are same and data across segments are different.

A. K Means Clustering:

K means or hard C means method is used to partition and analyse the data. K means partitions data in such a way that object inside each cluster remain closer to each other and away from objects in remaining cluster.

In K means, clustering algorithm starts with choosing number of required clusters K. Initially these numbers of clusters are taken as starting values. Then each point is examined and assigned to nearest cluster. When all the data points are assigned to cluster then new K centroids are recalculated [4].

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

K means algorithm s very simple but it does not give the same result for each run.

B. Fuzzy C Means Clustering:
For classifier design and feature analysis fuzzy C Means is widely used fuzzy clustering algorithm. Generalization of ISODATA gives FCM algorithm. The FCM algorithm finds partition for set of data points and minimizes objective function. For FCM sample set is given by,

$$X = \{x_1, x_2, ..., x_n\} \qquad (1)$$

where, n corresponds to number of data points and u_{ik} represents the membership degree of x_i for class i. U= u_{ik} is fuzzy classification matrix.
Objective function is given by equation (2) and it is iteratively minimized.

$$J(u,v) = \sum_{i=1}^{C}\sum_{k=1}^{n} u_{ik}^{m} \left| x_k - v_i \right|^2 \qquad (2)$$

Value of C is in between 2 and n [4], [5].
Cluster centre is calculated using

$$V_i = \frac{1}{\sum_{k=1}^{n} u_{ik}^{m}} \sum_{k=1}^{n} u_{ik}^{m} x_{ik} \quad i=1,2,....c \qquad (3)$$

Membership values are updated by,

$$u_{ik} = \frac{\left[\frac{1}{\left|x_k - v_i\right|^2}\right]^{1/m-1}}{\sum_{j=1}^{C}\left[\frac{1}{\left|x_k - v_j\right|^2}\right]^{1/m-1}} \qquad (4)$$

Once samples are grouped, cluster centres are needed to be recomputed in order to minimize j.

III. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is statistical learning based supervised classification system pioneered by Vapnik [6]. Structural Risk Minimization (SRM) principle is employed in SVM. For regression and classification problems SVM is very good tool [7]. It gives good generalization performance. The SVM is a linear machine that maximizes the margin by building a model for transforming low dimension feature space to high dimension feature space [8]. SVM finds applications in text categorization, time series analysis, database mining and face identification.

A. Linear Support Vector Machine:
SVM constructs hyperplane as decision plane in order to separate negative and positive classes with largest possible margin. For binary classification initially feature vector extraction is performed. Let $x_i \in R^d$ as training data with $y_i \in \{-1,+1\}$ i=1,2,……l for all training data [9]. Here l represents number of data and d corresponds to problem dimension. The hyperplane with largest possible margin is known optimal hyperplane and data points closed to optimal hyperplane are called support vectors [10], [11]. Optimal hyperplane and support vectors are shown in fig. 1.
Optimal hyperplane is given by,

$$w.x + b = 0 \qquad (5)$$

where, w is weigh vector and b is bias.

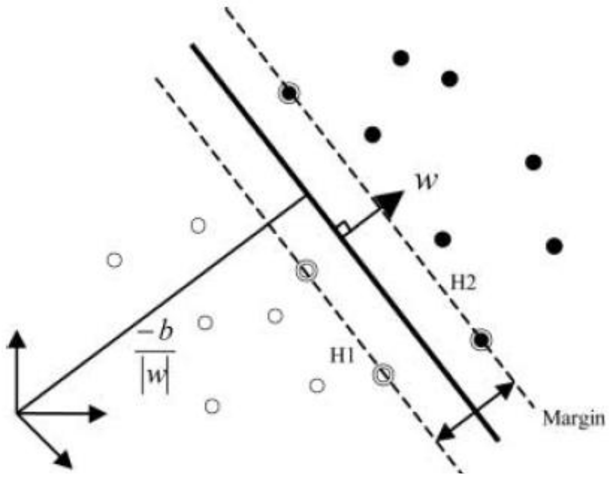

Fig. 1 Linear SVM. Support vectors are circled

B. Multi-Class Support Vector Machine:
Initially SVMs were developed for binary classification. But some classification problems contain more than two classes. So by combining large number of binary classifiers multi-class classifies can be obtained [10], [12].

IV. K NEAREST NEIGHBOR CLASSIFIER

KNN is the simplest classification technique. The KNN is used to classify the objects on the basis of most similar or closest training samples in the feature space. Majority vote of neighbors is used to classify the object. For an unknown sample and known training data, all the distances between all training set samples and unknown samples can be calculated. The smallest distance corresponds to training set sample close to unknown sample. So unknown sample can be classified on the basis of nearest neighbor [13].
The distance used to determine nearest neighbor is given by,

$$d(x,y) = \sqrt{(x-y)^T \sum^{-1} (x-y)} \qquad (6)$$

When K=1 it is called as nearest neighbor.

V. RESULTS

For this project, georeferenced .tiff (Tagged Image File Format) images, namely BAND2, BAND3 and BAND4, for different band data were used. BAND2 and BAND3 correspond to GREEN and RED in visible region and BAND4 corresponds to near-infrared region in the spectral band. BAND3 is used for training and BAND4 is used for testing.

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

In the proposed system adaptive histogram equalization is used to enhance the contrast of image. K means and Fuzzy C Means clustering techniques are used to obtain segmented images. Image is segmented and feature vector are calculated for all 3 regions. Classification is carried out using multi-class SVM and KNN.

Classification Results:

A. K means Clustering with SVM:

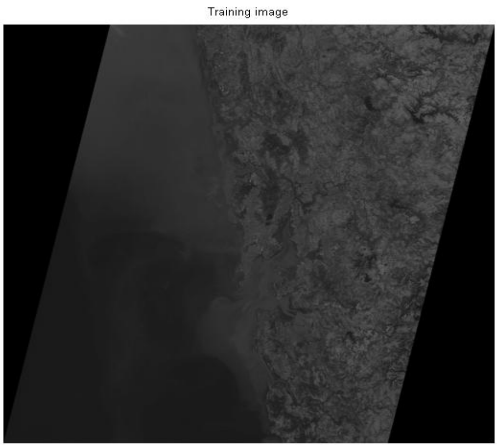

Fig. 2 Training Image

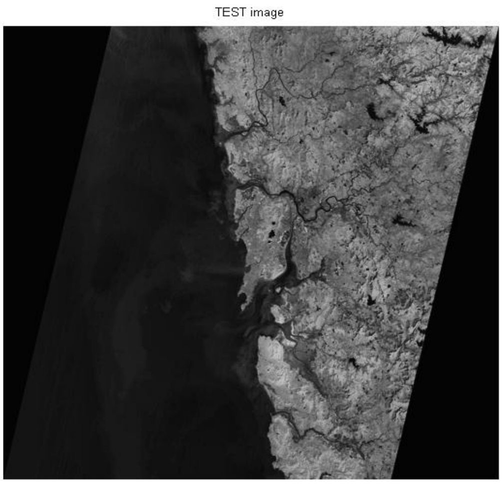

Fig. 3 Test Image

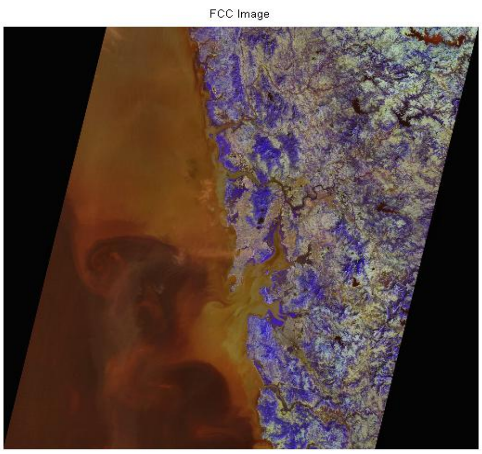

Fig. 4 FCC image

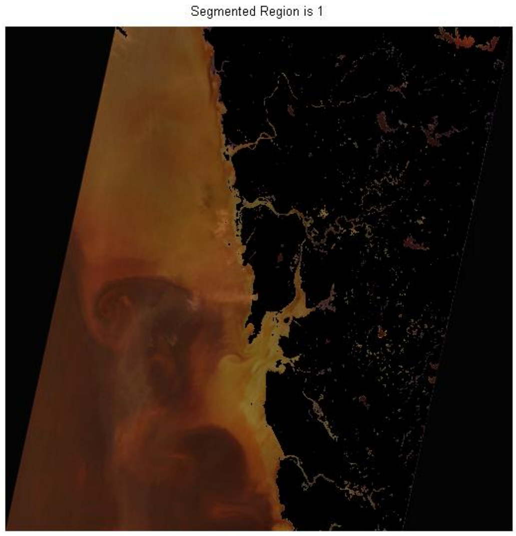

Fig. 5 Region 1

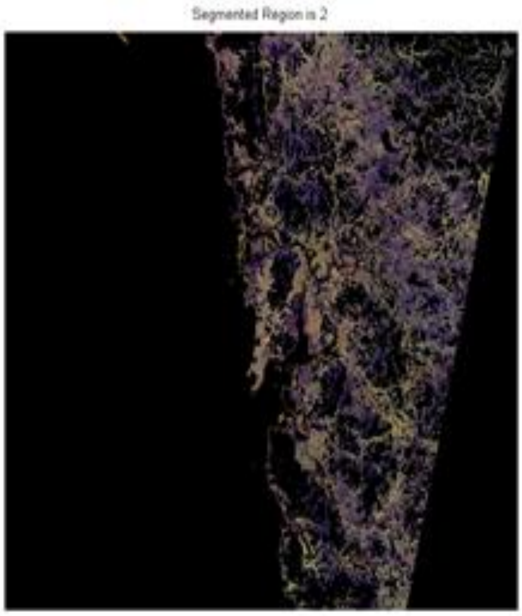

Fig. 6 Region 2

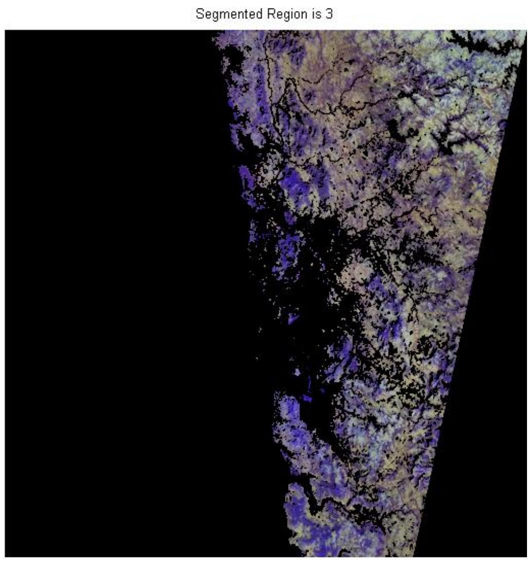

Fig. 7 Region 3

DOI 10.17148/IJIREEICE.2016.4637

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

B. K Means Clustering with KNN:

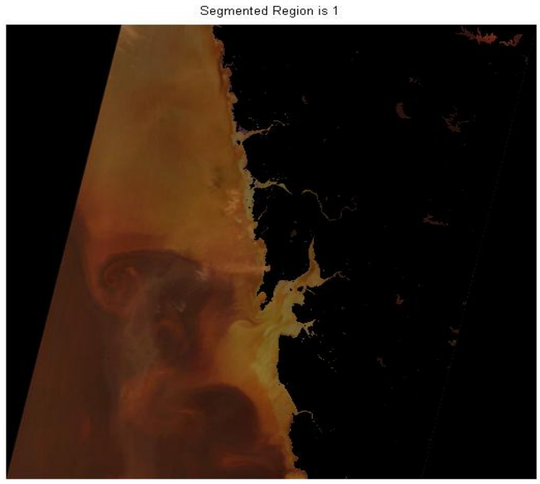

Fig. 8 Region 1

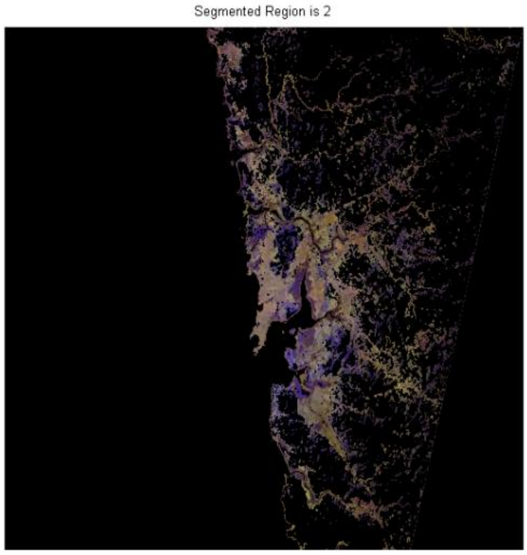

Fig. 9 Region 2

C. Fuzzy C Means Clustering with SVM:

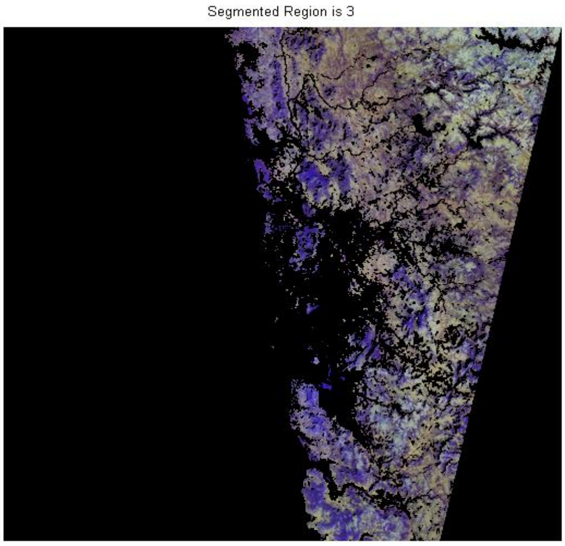

Fig. 10 Region 3

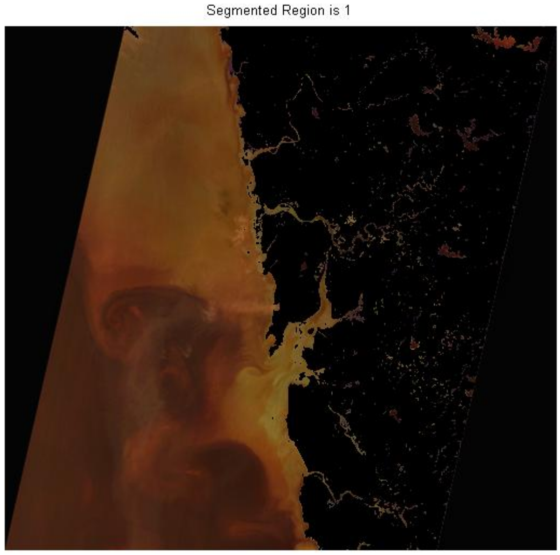

Fig. 11 Region 1

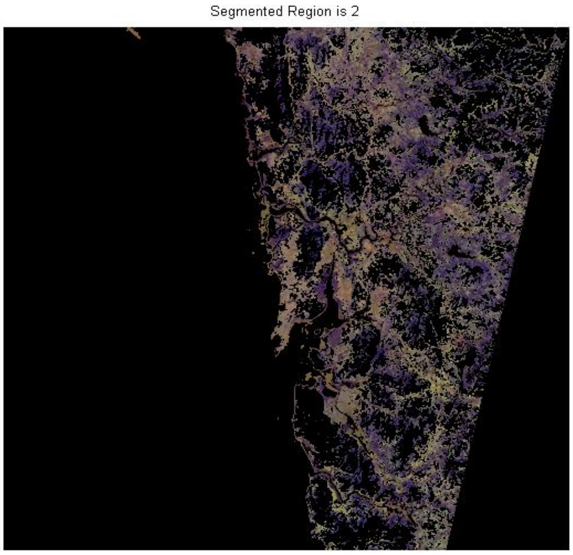

Fig. 12 Region 2

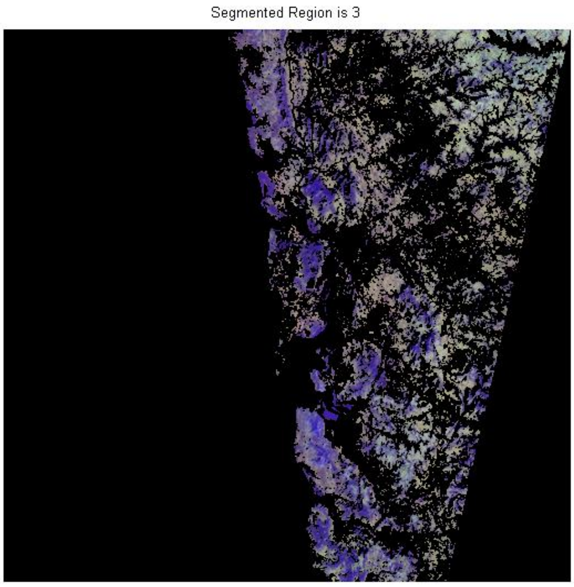

Fig. 13 Region 3

D. Fuzzy C Means Clustering with KNN:

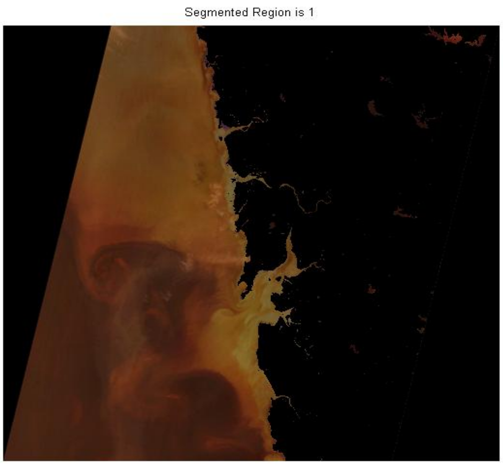

Fig. 14 Region 1

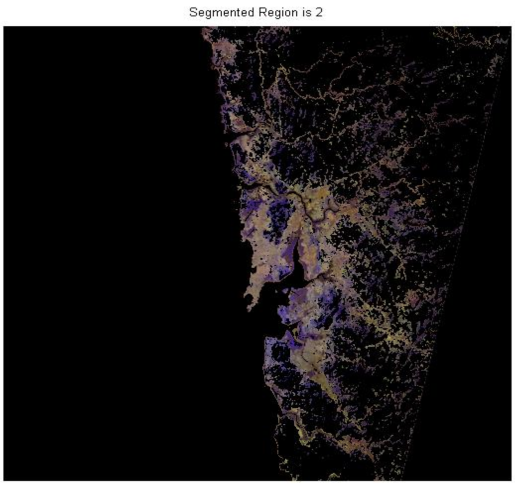

Fig. 15 Region2

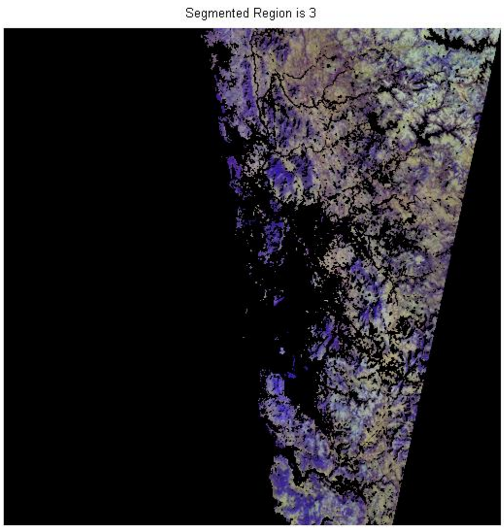

Fig. 16 Region 3

Fig. 2 is training image of BAND3 and Fig. 3 is test image of BAND4. Fig. 3 is FCC image obtained by combining BAND2, BAND3 and BAND4 images. Fig. 5, 6 and 7 represents three classified regions using K means clustering with SVM. Fig. 8, 9 and 10 represents classified outputs using K means clustering with KNN. The results obtained using Fuzzy c means with SVM are shown in fig. 11, 12 and 13. The classified outputs using Fuzzy c means with KNN are represented in fig 14, 15 and 16.

VI. CONCLUSION

This paper introduced a classification system for IRS satellite images using KNN and multi-class SVM. In proposed system image enhancement is carried out using adaptive histogram equalization. Segmented data is obtained using K means and FCM. Using multi-class SVM and KNN test image is classified into three classes namely water (region1), soil (region 2) and vegetation (region 3). From observation it is clear that FCM with multi-class SVM gives better separation between the three regions than KNN.

REFERENCES

[1] Pushpendra Singh Sisodia, Vivekanand Tiwari, Anil Kumar, "A Comparative Analysis of Remote Sensing Image Classification Techniques", 978-1-4799 3080-7/14/$31.00_c 2014 IEEE.

[2] C.Heltin Genitha, Dr.K.Vani, "Classification of Satellite Images using New Fuzzy Cluster Centroid for Unsupervised Classification Algorithm", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

[3] Omar S. Soliman, Amira S. Mahmoud, "A Classification System for Remote Sensing Satellite Images using Support Vector Machine with Non-Linear Kernel Functions", The 8th International Conference on INFOrmatics and Systems (INFOS2012) - 14-16 May.

[4] Soumi Ghosh, Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.

[5] S.V.S Prasad, Dr. T. Satya Savitri, Dr. I.V. Murali Krishna, "Classification of Multispectral Satellite Images using Clustering with SVM Classifier", International Journal of Computer Applications (0975 – 8887), Volume 35– No.5, December 2011.

[6] Ankur Dixit, Shefali Agarwal, "Comparison of Various Models and Optimum Range of its Parameters used in SVM Classification of Digital Satellite Image", Proceedings of the 2013 IEEE Second International Conference on Image Information Processing.

[7] Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar and Partha Pratim Sarkar, "A Tutorial on Different Classification Techniques for Remotely Sensed Imagery Datasets", Smart Computing Review, vol. 4, no. 1, February 2014.

[8] S.Manthira Moorthi, Indranil Misra, Rajdeep Kaur, Nikunj P Darji and R. Ramakrishnan, "Kernel based learning approach for satellite image classification using support vector machine", 978-1-4244-9477-4/11/$26.00 ©2011 IEEE.

[9] Wei WU, Guanglai GAO "Remote Sensing Image Classification with Multiple Classifiers based on Support Vector Machines", 2012 Fifth International Symposium on Computational Intelligence and Design.

[10] Falah Chamasemani, Yashwant Prasad Singh, "Multi-class Support Vector Machine (SVM) classifiers – An Application in Hypothyroid detection and Classification", 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications.

[11] Zhe Wang and Xiangyang Xue, "Multi-Class Support Vector Machine".

[12] Mahesh Pal, "Multiclass Approaches for Support Vector Machine Based Land Cover Classification".

[13] Sadegh Bafandeh Imandoust And Mohammad Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background", S B Imandoust et al. Int. Journal of Engineering Research and Applications, Vol. 3, Issue 5, Sep-Oct 2013, pp.605-610.

[14] Kanika Kalra, Anil Kumar Goswami, Rhythm Gupta "A Comparative Study of Supervised Image Classification Algorithms for Satellite Images", International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084 Volume-1, Issue-10, Dec-2013.