ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

Multi Spectral Image Classification and Quality Parameters using Random Forest Classifier

[1] K. Radhika, [2] S. Varadarajan
[1] ECE Dept, JNTUK, Kakinada, AP, INDIA [2] Secretary, APSCHE, AP, INDIA

Abstract: **-- Remote sensing is the art of obtaining information regarding an object or area using machine or device which is not physically contacted with the area. Geology, urban planning, Forest and land cover/land use are the several applications of remote sensing. Remote sensing is majorly utilized for generation of classification map. Latest methods used for classification of pixels in multispectral satellite images consists supervised classifiers such as the maximum-likelihood classifier, neural network classifiers, fuzzy neural networks, support vector machines ,k-NN and decision trees (Random Forest). SVM may be one-class or multi-class SVM. KNN is simple technique. In case of Random Forest, many decision trees are grown by it for classification. The input vector needs to run through every decision tree in the forest to classify a new object. The forest chooses the classification having the most votes. Random Forest provides a robust algorithm for classifying large datasets. The potential of Random Forest is not been explored in analyzing multispectral satellite images. KNN is simple technique in high-dimensional feature space. In case of Random Forest, many decision trees are grown by it for classification. The input vector needs to run through every decision tree in the forest to classify a new object. The forest chooses the classification having the most votes. Random Forest provides a robust algorithm for classifying large datasets. The potential of Random Forest is not been explored in analyzing multispectral satellite images.**

Keywords**- Classification, Multispectral, classification, satellite, Random Forest.**

I. INTRODUCTION

Multispectral images are consistently dissected by ordinary factual strategies, and delicate processing methods like neural systems, fuzzy interface systems and fuzzy neural systems. Customary strategies utilized for the classification of pixels in multispectral images comprises of the maximum likelihood classifier, the minimum distance classifier, and a few clustering techniques, for example, isodata. In maximum likelihood classicification, each pixel is inspected for every conceivable class and the pixel that is relegated to the class with the outrageous back likelihood. If there should arise an occurrence of neural systems, the information perception vector can be straightforwardly mapped to yield class, once a neural system is trained. In this way, for huge images neural systems are very appropriate.The examination have made by Huang and Lippmann [1] between neural systems and routine classifiers.

Support Vector Machine (SVM) was initially created in 1992, presented by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) are an arrangement of related supervised learning systems connected for grouping. There are various productions enumerating the scientific detailing and calculation improvement of the SVM [2], [3].

Classification in SVM is a case of Supervised Learning. Known lables will demonstrate whether thesystem is working appropriately or not. This data will coordinates to a required reaction, approves the precision of the system, or be utilizesd to help the system to learn out how to act correctly.SVM performs include determination or highlight extraction. At the point when expectation of obscure examples is redundant, highlight determination and SVM arrangement together can be utilized. They can be used to recognize key sets.

To classify objects k-closest neighbor [4] utilizes an instance-based learning procedure based with respect to the guideline of nearest preparing cases in the element space. It is very basic strategy to process. Obscure example can be arranged accurately by contrasting separation with deference with known specimens. Along these lines, obscure specimens are distinguished in view of closest neighbors.

Trees speak to another gathering of characterization calculations.Decision tree classifiers have not been utilized generally by the remote detecting group in spite of their non-parametric nature and their alluring properties of effortlessness in taking care of the non-ordinary, non-homogeneous and uproarious information [5].Decision tree classifiers are more effective than single-stage classifiers. With a choice tree classifier, choices are made at various levels.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

Decision tree classifiers are otherwise called multilevel classifiers. The fundamental worries in a Decision tree classifier are the detachment of gatherings at each non-terminal hub and the selection of components that are best in isolating the gathering of classes. In planning a Decision tree classifier it is alluring to develop an ideal tree in order to accomplish the most astounding conceivable grouping exactness with the base number of counts [6]. The double tree classifier is viewed as a unique instance of a choice tree classifier.

The Random Forest calculation has been utilized as a part of numerous information mining applications, be that as it may, its potential is not completely investigated for breaking down remotely detected pictures. Arbitrary Forest depends on tree classifiers. Irregular Forest develops numerous order trees. To group another component vector, the info vector is arranged with each of trees in the woods. Each tree gives a classification, and we say that the tree "votes" for that class. The timberland picks the arrangement having the most votes over every one of the trees in the woods. Among many favorable circumstances of Random Forest the noteworthy ones are: unexcelled exactness among current calculations, proficient execution on substantial informational indexes, and an effectively spared structure for later utilization of pre-created trees [5].

II CLASSIFICATION TECHNIQUES

A. One Class SVM:

Support Vector Machine (SVM) is measurable learning based directed order framework spearheaded by Vapnik [7]. The guideline utilized in SVM is Structural Risk Minimization (SRM). SVM is a decent device for relapse and characterization issues [8]. It displays great speculation execution. The SVM is a straight machine that builds the edge by building up a model for changing low measurement highlight space to high measurement include space [9].

In Linear Support Vector Machine, SVM bluids hyperplane as decision plane in order to distinguish negative and positive classes with largest possible margin. For binary classification priory 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 [10]. Here l denotes number of data and d related to problem dimension. The hyperplane with maximum possible margin is called optimal hyperplane and data points closely related to optimal hyperplane are called support vectors [11], [12].Optimal hyperplane can be derived from equation (1). Optimal hyperplane and support vectors are shown in fig.1.

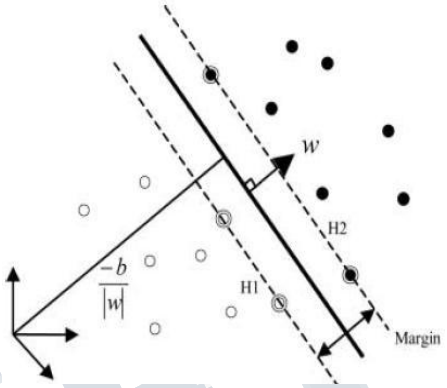

Fig. 1: Linear SVM. Support vectors are circled

Optimal hyper-plane is given by,

$$w.\ x+b = 0 \qquad (1)$$

Where, w is weigh vector and b is bias.

B. ***Multi-classification SVM*****:** Generally SVM is defined for two-class classification, but two possibilities are there to enhance the two-classification SVM to a multi-classification SVM. However, because of the high dimensional matrix and complex computation, this approach is undesirable. The other one is to build a multi-classification classifier by a few two-classification classifiers. The generally used multi-classification classifiers are described as follows:

(1) One-against-rest SVM

One-against-rest SVM consists of K two-class classification classifiers to obtain K-class classification. When the number of classes is large, classification is effected by the unbalance of the training samples.

(2) One-against-one SVM

One-against-one SVM use K(K-l)/2 two-class classification classifiers to obtain K-class classification, which is better than one-against-rest SVM, and has the good promoting ability. However, it also suffers from the problem of large computation and error accumulation.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

(3) Hierarchical SVM

Hierarchical SVM use (K-I) two-class classification classifiers to obtain a K-class classification. In this method, there is no sub-regional and the whole number of classifiers is less than one-against-one SVM .these become advantageous. However, it also suffers from the problem of error accumulation and promotion.

(4) Directed acyclic graph (DAG)

DAG constructs a multi-class classifier by few two-class classifiers having the similar training processing with one against-one SVM. DAG consists K(K-l)/2 two-class classifiers, which utilizes the guidance from the root node acyclic graph (Fig. 2). In this case, error only depends on the class number K and the intervals of the node, and it has no relationship with the dimension of the input space.

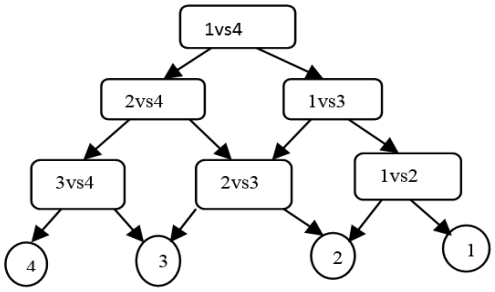

Fig. 2: Directed acyclic graph

Because the high dimensional feature vectors can affect the speed of SVM division, and DAG has no relation with the dimension of the input space, and comparing with the above advantages of multi-class classifiers, we use DAG as the multi-class classifier in the following experiment.

C. K-Nearest Neighbors: To classify objects k-nearest neighbor [4] uses an instance-based learning technique based on the principle of closest training examples in the feature space. It is quite simple technique to compute. Unknown sample can be classified correctly by comparing distance with respect to known samples. So, unknown samples are identified based on nearest neighbors. The smallest distance corresponding unknown and known sample will decide similar instances. For every data point, which is in the k neighboring training data, the points that are nearest in distance to that data point are selected. Based on certain distance metric, the class that comes most of times in the neighbourhood is assigned to new data point. A simple Euclidean distance (for continuous variables) or the Hamming distance (for discrete variables)is used as distance metric in this approach. It is also called a lazy leaner because, all calculation are delayed until classification. It is most useful where the data exhibit spatial properties in order to perform classification.

KNN is the simplest classification technique on the basis of most similar or closest training samples in the feature space. The k-NN classifier is solely on the principle of learning by example, that is, a given test instance is compared with training instances that are nearer to it. The training instances, or training tuples, are denoted by n attributes. Every instance indicates a data point in the entire n-D pattern space or feature space. The k-NN classifier searches in the feature space for finding the k training instances nearest to a given query-instance or unknown tuple. In this way the obtained k training tuples are definitely the k —nearest neighbors of the examined query-instance. The term —nearest can be defined in terms of Euclidean distance as the distance metric.

Considering a simple k-NN example for k = 7 and a given query-instance xq as shown in Figure 3. At first, after choosing the value of k, need to compute the distance among the query-instance and all training tuples based on distance metric. Nearest neighbour k can be find by sorting the distances for all training tuples nearest neighbor to k—the minimum distance. Later we need to examine all the sets of the training data for the sorted value that come under k. In this example shown, the query-instance x_q is classified as '–' (negative) since 4 of its nearest neighbors are indicated as '–'. This procedure for classification is rely solely on majority voting by its nearest neighbors.

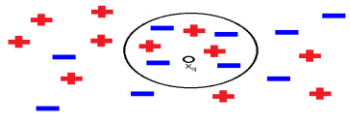

Fig 3: A k-NN example with k=7

D. Random Forest (RF): Up to now Random Forest algorithm has been used in various data mining applications but its potential is not entirely used for remotely sensed

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

images. Random Forest is based on tree classifiers. Many classification trees are grown by Random Forest. The input vector is classified with each of trees in the forest to classify a new feature vector. Every tree gives a classification, and this tree "votes" for that class. The tree which acquires majority votes will be selected for classification by the forest. Among various advantages of Random Forest the major ones are: increased accuracy among current algorithms, efficient implementation on large data sets, and an easily saved structure for future use of pre-generated trees [11].

Random forest classifier performs similarly as SVMs in terms of classification accuracy and training time. The number of user-defined parameters needed by random forest classifiers is less than the number needed for SVMs and simpler to define. The general technique used in random forest is bootstrap aggregating, or bagging, to tree learners. Let a set $X = x_1, ..., x_n$ with responses $Y= y_1, ..., y_n$, bagging iteratively (B times) selects a random sample with replacement of the training set and fits trees to these samples:

For b = 1, ..., B:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b.
2. Train a decision or regression tree f_b on X_b, Y_b.

After training there are some unseen samples x'. Predictions for unseen samples x' can be carried out by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x') \qquad (2)$$

or by taking the majority vote in the case of decision trees.

In bootstrapping procedure the variance of the model decreases without increasing the bias, which leads to better model performance. This means that while the predictions of a single tree are more sensitive to noise in its training set, the average of several trees is not, as long as the trees are not correlated. Strongly correlated trees can be obtained by simply training several trees on a single training set. Bootstrap sampling, de-correlates the trees by showing them various training sets. The number of samples/trees, B, is a free parameter. Typically several number of trees are used based on size and nature of the training set. A minimum number of trees B can be found using cross-validation, or by seeing the out-of-bag error: the mean prediction error on every training sample x_i, using purely the trees that are not having x_i in their bootstrap sample [13]. The training and test error tend to level off after few number of trees have been fit.

In the procedure discussed up to now ,the original bagging algorithm for trees. The difference between Random forests and general scheme is they use a modified tree learning algorithm process, a random subset of the features. This process is known as "feature bagging". The main reason to do this is the correlation among trees in an ordinary bootstrap sample: if one or a few features are very strong predictors for the response variable (target output), these features will be selected in many of the B trees, causing them to become correlated. An analysis of how bagging and random subspace projection contribute to accuracy gains under different conditions is given by Ho [14].

Randomized trees, or Extra Trees can be obtained by randomization. These extra trees are trained by bagging and the random subspace method, such as in an ordinary random forest, the top-down splitting in the tree learner is randomized. A random value is selected instead of computing split combination (based on, e.g., information gain or the Gini impurity), for each feature under consideration. This value is selected from the feature's empirical range (in the tree's training set, i.e., the bootstrap sample).

III. RESULTS AND CONCLUSION

In this paper different classifiers have been analysed. The classified image and ground truth data has been verified in validation process. 180 ground truth points have been considered for validation purpose. More number of points lead to accurate measurement. Validation process gives confusion matrix of four or five classes [15-16]. The quality parameters like overall accuracy and producer's accuracy, user's accuracy, precision, recall, specificity and f1 score of all the classes and overall classes have been measured.

Figure 4 shows the comparison of accuracy measurement of methods SVM, K-NN and random forest. Figure 5 shows the comparison of precision measurement of methods SVM, K-NN and random forest. Figure 6 shows the comparison of recall measurement of methods SVM, K-NN and random forest. Figure 7 shows the comparison of specificity measurement of methods SVM, K-NN and random forest.

Confusion matrix, accuracy measurements and quality

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

parameters have been evaluated for methods SVM, K-NN and random forest. Only random forest analysis have been given in this paper.

Table 1 shows the Confusion matrix for RF method. 4 classes have been considered for this confusion matrix. Table 2 shows the accuracy measurements like producer's accuracy, user's accuracy of all classes and overall accuracy. Table 3 shows quality parameters for the confusion matrix shown in table 1.

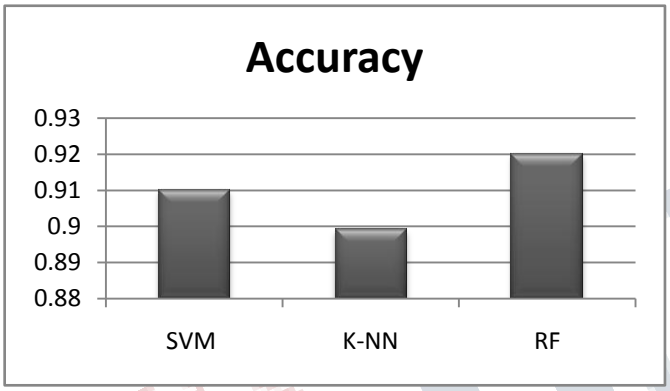

Fig 4: Comparison of accuracy measurement

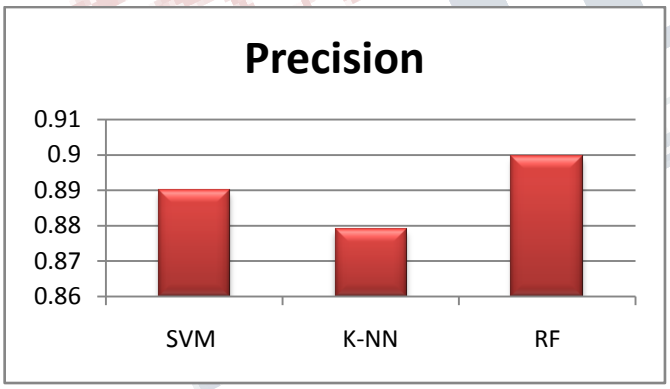

Fig 5: Comparison of precision measurement

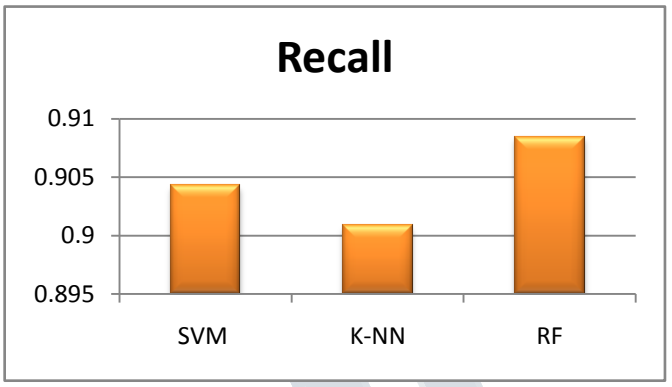

Fig 6: Comparison of recall measurement

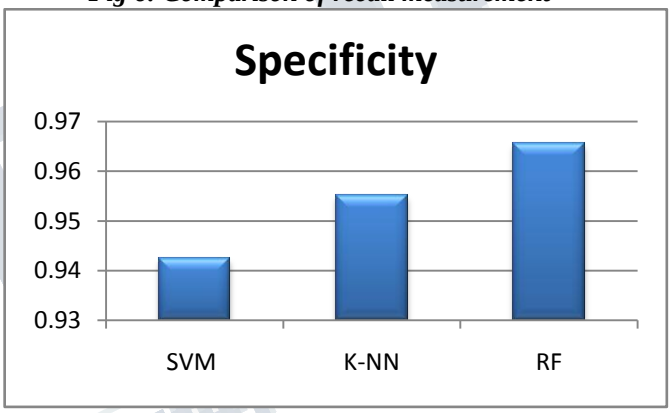

Fig 7: Comparison of specificity measurement

Table 1: Confusion matrix for RF method

PRIDICTED					
CLASS	**A**	**B**	**C**	**D**	**Total**
A	**36**	1	4	3	44
B	0	**33**	1	1	35
C	2	4	**46**	1	53
D	4	2	0	**42**	48
Total	42	40	51	47	180

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 3, March 2017

Table 2: Accuracy measurement for RF method

Type of land cover	Reference Pixels	Classified Pixels	Matching	Accuracy type	
				Procedures	Users
A	42	44	36	85.71%	81.82%
B	40	35	33	82.50%	94.29%
C	51	53	46	90.20%	86.79%
D	47	48	42	89.36%	87.50%
Total	180	180	157		
Overall Classification Accuracy				87.22222222	

Table 3: Quality parameters for RF method

	Accuracy	**Precision**	**Recall**	**Specificity**	**F1 score**
A		0.8571	0.8181	0.9558	0.8372
B		0.825	0.9428	0.9517	0.88
C		0.9019	0.8679	0.9606	0.8846
D		0.8936	0.875	0.9621	0.8842
Overall	0.8722	0.8694	0.8759	0.9575	0.8715

Acknowledgment

The authors would like to thank Indian Space Research Organization (ISRO) and officials related to USGS explorer site and Google earth site for providing satellite data.

REFERENCES

[1] W. Y. Huang and R. P. Lippmann, "Neural Net and Traditional Classifiers," in Neural Information Processing Systems, 1988, pp. 387–396.

[2] V.N.Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.

[3] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, 2 (2), pp. 121–167, 1998.

[4] G. Shakhnarovich, T. Darrell, P. Indyk, ―Nearest neighbor methods in learning and vision: Theory and Practice,‖ MIT Press, 2005.

[5] M. Ghose, R. Pradhan, and S. Ghose, "Decision tree classification of remotely sensed satellite data using spectral separability matrix," International Journal of Advanced Computer Science and Applications, vol. 1, no. 5, pp. 93–101, 2010.

[6] A. D. Kulkarni, Computer Vision and Fuzzy Neural Systems. Upper Saddle River, NJ: Prentice Hall,2001.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Ryad Malik, Radja Kheddam and Aichouche Belhadj-

Aissa.," Object-Oriented SVM Classifier for ALSAT-2A High Spatial Resolution Imagery: A Case Study of Algiers Urban Area", 978-1-4799-8637-8/15/$31.00 ©2015 IEEE.

[13] M. Ghose, R. Pradhan, and S. Ghose, "Decision tree classification of remotely sensed satellite data using spectral separability matrix," International Journal of Advanced Computer Science and Applications, vol. 1, no. 5, pp. 93–101, 2010.

[14] A. D. Kulkarni, Computer Vision and Fuzzy Neural Systems. Upper Saddle River, NJ: Prentice Hall, 2001.

[15] Y. Murali Mohan Babu, M.V. Subramanyam & M.N. Giriprasad, "Fusion and Texture based Classification of Indian Microwave Data - A Comparative Study", International Journal of Applied Engineering Research, Volume 10, Issue 1, 1003-1010, February 2015.

[16] M.V. Subramanyam & M.N. Giriprasad, "Fusion based land cover classification of Risat-1 data", in an International Conference on Industrial Electrical and Electronics and Electronics Engineering (ICIEEE), Bangkok, August 2014.