International Journal of Innovations in Engineering and Education
www.ijiee.com Volume-01, Issue-01

Image Classification based on Ensemble Subspace Discriminant method using Random Subspace Algorithm

[*1]K. Radhika and [2]D.NagaJyothi

[1,2] ECE Dept., GIST, Nellore, AP, INDIA

Abstract-Classification of images obtained from satellites is a challenge now-a-days in order to know the information contained by them. The ensemble can be used for weak learners also. To judge whether the classifier is giving correct information or not validation is important. To validate this ensemble technique we have considered 50,100,150 and 200 sample points, examined the accuracy and necessary parameters. The nature of the proposed method for the above said number of sample points is discussed.

Keywords - Classification; Satellite; Image; Random subspace; Ensemble.

1. INTRODUCTION

Classification is playing major role in the extraction of information contained by the satellite images. During this process of classification several problems may come across like accuracy, misclassification etc. There are several classifiers like Support vector machine, K nearest neighbors, neural network, maximum-likelihood [1-3]. But to improve accuracy and to reduce misclassification single classifier is not sufficient [4].

The prediction capability of the ensemble classifier model can be increased by the attributes. If we repeated N times, N number of new data sets will be generated. After this, boosting is required to enhance the quality of the ensemble model. Random subspace is the majorly using ensemble strategy and this procedure is taken place in feature space. This will combine a set of weak classifiers into a strong classifier.

2. ENSEMBLE BASED CLASSIFIERS

This can be possible by combining several methods. These are developed to increase the performance of the prediction. All the benefits of individual classifier can be reached with ensemble methodology.

2.1 Random Subspace Classifier

This is one kind of ensemble classifier. This will have several classifiers in a subspace of data feature space. The output of this classifier is depends on individual classifier results by majority voting rule. The classifiers that can be combined to get Random subspace classifier are linear classifier, nearest-neighbour classification, support vector machine, etc.

2.2 RS Algorithm

Random Subspace will build the one base classifier on each subset by taking samples of original feature set.

Algorithm:

Input training data

U_m is the base classifier, m=1,2,.....,M

Initialize the weight distribution of training data $w_i=1/n, i=1,2,,.....n$

For m= 1 to M

Compute the misclassification error rate $\varepsilon_m=P(T_m(xi)\neq yi)$

Calculate coefficient of U_m , $U = sign\sum_{k=1}^{K} U_m$

$$a_m = \frac{1}{2}\log\frac{1-\varepsilon_m}{\varepsilon_m}$$

Normalization constant is $Z_m = \sum_{i=1}^{N} w_i \exp\left(-a_m y_i U_m\left(x_i\right)\right)$

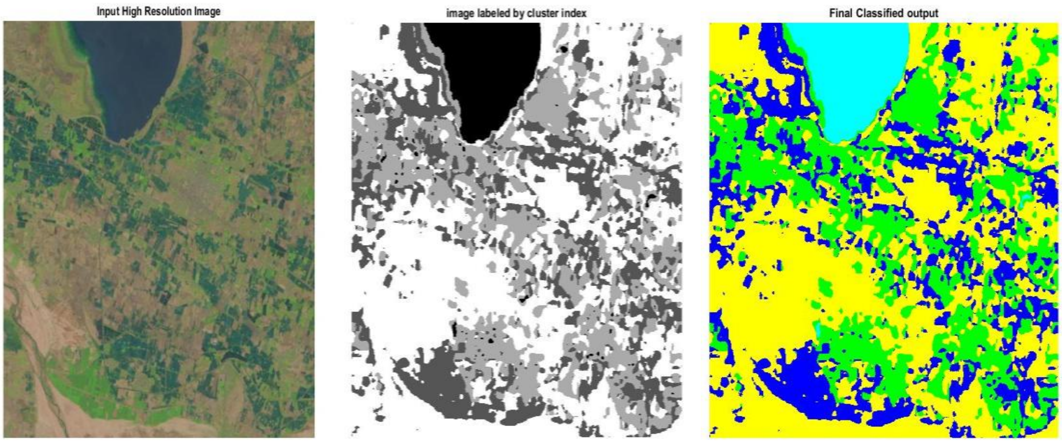
Update the weight $w_i = \frac{w_i}{z_m}\exp\left(-a_m y_i U_m\left(x_i\right)\right)$

End

Output $U\left(x\right) = sign\left(\sum_{m=1}^{M} a_m U_m\left(x\right)\right)$

3. RESULTS AND DISCUSSIONS

The classes defined here as A, B, C, and D. The optical or satellite input image is shown in figure 1(a). Figure 1(b) is the clustered output with SOM algorithm. The entire procedure here is of two parts, first one is clustering process and last one is classification process. The clustered output of given satellite image is shown in figure 1(b). Figure 1(b) is given to ensemble classifier. The classifier output is shown in figure 1(c). Table 1, is the confusion matrix of validation with 50 ground truth points. And Table 2 is the quality parameter measurements. A table 3 is the confusion matrix of 100 validation points. A table 4 is the different parameters along with overall accuracy. Observing all the tables, there no much change in overall accuracy of different number of validation points. The table 5 and table 7 are the confusion matrix of 150 and 200 validation points. The table 6 and table 8 are the overall accuracy of different number of validation points.



(a) (b) (c)

Figure 1: Input image, clustered image and output image

Table 1: Confusion matrix with 50 validation points

		PRIDICTED				
ACTUAL	CLASS	A	B	C	D	Total
	A	10	2	0	0	12
	B	1	13	0	0	14
	C	0	0	15	0	15
	D	0	0	0	9	9
	Total	11	15	15	9	50

Table 2: Quality measurements with 50 validation points

	Accuracy	Precision	Recall	Specificity	F1 score
A	0.94	0.909091	0.833333	0.973684	0.869565
B		0.866667	0.928571	0.944444	0.896552
C		1	1	1	1
D		1	1	1	1
Over all		0.943939	0.940476	0.979532	0.941529

Table 3: Confusion matrix with 100 validation points

		PRIDICTED				
ACTUAL	CLASS	A	B	C	D	Total
	A	18	3	0	2	23
	B	2	34	0	0	36
	C	0	0	31	0	31
	D	0	0	0	10	10
	Total	20	37	31	12	100

Table 4: Quality measurements with 100 validation points

	Accuracy	Precision	Recall	Specificity	F1 score
A	0.93	0.9	0.782609	0.974026	0.837209
B		0.918919	0.944444	0.953125	0.931507
C		1	1	1	1
D		0.833333	1	0.977778	0.909091
Over all		0.913063	0.931763	0.976232	0.919452

Table 5: Confusion matrix with 150 validation points

		PRIDICTED				
ACTUAL	CLASS	A	B	C	D	Total
	A	33	0	2	1	36
	B	2	42	1	2	47
	C	2	2	33	0	37
	D	0	2	0	28	30
	Total	37	46	36	31	150

Table 6: Quality measurements with 150 validation points

	Accuracy	Precision	Recall	Specificity	F1 score
A	0.9067	0.8919	0.9167	0.9649	0.9041
B		0.9130	0.8936	0.9612	0.9032
C		0.9167	0.8919	0.9735	0.9041
D		0.9032	0.9333	0.9750	0.9180
Over all		0.9062	0.9089	0.9686	0.9074

Table 7: Confusion matrix with 200 validation points

		PRIDICTED				
ACTUAL	CLASS	A	B	C	D	Total
	A	44	0	2	1	47
	B	2	60	1	2	65
	C	2	2	42	0	46
	D	0	2	0	40	42
	Total	48	64	45	43	200

Table 8: Quality measurements with 200 validation points

	Accuracy	Precision	Recall	Specificity	F1 score
A	0.93	0.9167	0.9362	0.9739	0.9263
B		0.9375	0.9231	0.9704	0.9302
C		0.9333	0.9130	0.9805	0.9231
D		0.9302	0.9524	0.9810	0.9412
Over all		0.9294	0.9312	0.9764	0.9302

4. CONCLUSION

The algorithm is checked for 50, 100, 150 and 200 ground truth points and the corresponding overall accuracies and different quality parameters are calculated. The overall accuracy values are obtained are all similar almost in each case. The theme of usage of the different points was to check how far the values are accurate. All the four sets have given similar values in the accuracies and different quality parameters. Generally, with less number of validation points the judgement of a procedure is not correct. In this connection we have verified for different set of validation points. So, it can be concluded that our validation points are correct to use for further procedures.

REFERENCES

[1] S.Nithyakalyani,and B. Gopinath;"Analysis of Node Clustering Algorithms on Data Aggregation in Wireless Sensor Network", J SciInd Res,vol. 75, 38-42, 2016.
[2] S J. Lee,T R Lin Grace,and P H. His;"Industrial Cluster Development and its Contribution to Economic Growth in Taiwan - Hsinchu Science and Industrial Park (HSIP)", J SciInd Res, Vol.76, 273-278, 2017.
[3] K. Netti,and Y. Radhika;"A Model for Accurate Prediction in GeoRSS Data Using Naive Bayes Classifier", J SciInd Res, vol. 76, 473-476,2017.
[4] S A. Mousavi, Z. Hanifeloo, P. Sumari, and MRM. Arshad;"Enhancing the Diagnosis of Corn Pests using Gabor Wavelet Features and SVM Classification", J SciInd Res, 75, 349-354, 2016.