EVALUATION OF MULTIPLE CLASSIFIER COMBINATION TECHNIQUES FOR LAND COVER CLASSIFICATION USING MULTISOURCE REMOTE SENSING DATA

Hai Tung CHU, Linlin GE and Rattanasuda CHOLATHAT

School of Surveying and Geospatial Engineering,
The University of New South Wales, Sydney, NSW 2052 Australia
Email: chuhaitung@gmail.com; ht.chu@student.unsw.edu.au

KEY WORDS: multi-source remote sensing data, multiple classification systems, Artificial Neural Network, Support Vector Machine, Self-Organizing Map

Abstract: Use of multisource remote sensing data, particularly Synthetic Aperture Radar (SAR) and optical images, can improve performance of land cover classification since many types of information are involved in the classification process. Recently, the multiple classification systems (MCS) or ensemble classifiers has been developed and increasingly used for classifying remote sensing data. It is considered as a promising approach to increase the classification accuracy.
In this paper, different classification combination methods were carried out and evaluated for classifying land cover features in New South Wales, Australia using various integrations of SAR (ALOS/PALSAR, ENVISAT/ASAR) and optical (Landsat 5 TM+) satellite images and their derivative products such as textural information, Normalized Difference Vegetation Index (NDVI). Three classifiers were applied for classification, including Artificial Neural Network (ANN), Support Vector Machine and Self-Organizing Map (SOM). The outputs of these classifiers were then fused using various combination schemes such as majority voting, sum, evidence reasoning (Dempster-Shafer) theory. The other approach involve using other well known MCS techniques, namely, bagging and boosting algorithms were also carried out for each of the classifier.
Results of the study illustrated the advantages of the multiple classification combination approach for land cover classifications, especially in conjunction with multisource remote sensing data. In most of cases, the multiple classifier system outperformed the single classifier and gave a noticeable improvement in the classification accuracy. The experiments also revealed that the multisource datasets always gave better accuracy than that obtained by the single-source datasets.

1. INTRODUCTION

The use of multisource remote sensing data, particularly SAR coupled with optical images, hold potentials to improve land cover classification performance because of the complementary characteristic from different kind of data can contribute to the improvement of the classification performance. This approach has been fueled by an availability of a large variety of satellite imagery and become very attractive to researchers. Many studies have been carried out using combinations of different kinds of remote sensing data and have showed very positive results (e.g. Chust et al. 2004, Erasmi & Twelve 2009, Chu & Ge 2010a, Soria-Ruiz et al. 2010). For example, Erasmi & Twele (2009) used dual-polarimetric SAR (Envisat/ASAR) satellite image data and optical medium resolution (Landsat ETM+) data for classifying land cover features at the regional level in Central Sulawesi, Indonesia. The authors pointed out that the integration of ASAR with Landsat images increased classification accuracy significantly, and the combination of like-polarised ASAR time series and Landsat multi-spectral data produced the best results. Lehmann et al. (2011) combined Landsat TM and ALOS/PALSAR data for forest monitoring in north-eastern Tasmania, Australia. The results indicate that the synergistic use of SAR and optical data produced a better forest classification than that of either single dataset. The combination approach provided the highest classification accuracy of 94.32%, with 5.7% and 2.3% improvement in overall accuracy over the PALSAR-only and Landsat-only classifications, respectively.

The classification algorithms are also very important for the success of the land cover classification process. A large range of classification algorithms has been developed and applied for classifying remotely sensed data. The traditional parametric classifiers such as the Minimum Distance, Maximum Likelihood (ML) classifiers have been used extensively (Waske & Braun, 2009) due to its acceptable accuracy and fast performance. However, the major limitation of the parametric algorithms is their reliance on the assumption of normal distribution of input data – which is often not true for remotely sensed data (Waske & Benediktson, 2007). This limitation makes it difficult for such parametric classifiers to handle complex datasets consisting of different kind of data such as multisource data. On the other hand, non-parametric classifiers such as the Artificial Neural Network (ANN), Support Vector Machine (SVM) or Self-Organizing Map does not constrain to the assumption of normal distribution, and are therefore often considered more appropriate for classifying remotely sensed data. Many studies have been conducted using these non-parametric algorithms to classify remote sensing data (Kavzoglu and Mather 2003, Dixon and Candade 2008, Salah et al. 2009, Waske and Benediksson 2007, Kavzoglu and Colkesen 2009). Although results vary considerably, most authors reported that the non-parametric classifier is robust and often outperform the traditional parametric classifier.

Another powerful technique for mapping land cover features is Multiple Classifier System (MCS) or classifier ensemble. MCS is capable to integrate advantages and alleviate weaknesses of constituent classifiers. Furthermore, the MCS allows minimize the risk of poor selection (Polikar, 2006). MCS involves different classification strategies such as parallel or hierarchical computing, Bagging and Boosting, and different classifier combination rules, such as Majority Voting, Sum, Max, Min, Product, Fuzzy integral or evidence reasoning based on Dempster-Shafer theory. Some basic concepts and application of MCS for classifying remote sensing data was discussed in detail by Benediksson et al. (2007). Tzeng et al. (2006) applied MCS with the bagging and boosting algorithms to classify land cover features in Taiwan using multispectral data similar to SPOT satellite images. Results showed that the MCS improved the classification accuracy compared to the single classifier. Salah et al. (2010) employed the Fuzzy Majority Voting techniques to combine classification results of three classifiers over four different study areas using Lidar and aerial images. The results illustrated that while the overall classification accuracies were slightly improved, the commission and omission errors were reduced considerably compared to the best individual classifier.

In this paper, the capability of the MCS techniques is investigated for classifying mutltisource remote sensing data using different non-parametric constituent classifiers.

2. STUDY AREA AND DATA USES

The study area is located in Appin, in the state of New South Wales, Australia. The centre coordinate is 150^{o} 44' 30" E; 34^{o} 12' 30" S. The study area is mixed by different kinds of land cover features, including native dense forest, grazing land, urban and rural residential areas, facilities and water surfaces.

Three types of remote sensing data were employed for this study. These include:

SAR: Six ENVISAT/ASAR VV polarisation and six ALOS/PALSAR HH polarisation images acquired in 2010 (Figure 1, 2 and Table 1).

Optical: Three Landsat 5 TM images acquired on 25/03/2010, 10/9/2010 (Figure 3) and 31/12/2010 with seven spectral bands and spatial resolution of 30m. In this study six spectral bands (except the thermal band) were used.

Table 1: ENVISAT/ASAR and ALOS/PALSAR images for the study area.

Satellite/Sensors	Date	Polarisation	Mode
ENVISAT/ASAR	03/04/2010	VV	Descending
	24/06/2010	VV	Ascending
	25/06/2010	VV	Descending
	27/06/2010	VV	Ascending
	28/06/2010	VV	Descending
	25/09/2010	VV	Descending
ALOS/PALSAR	04/01/2010	HH	Ascending
	22/05/2010	HH	Ascending
	07/07/2010	HH	Ascending
	22/08/2010	HH	Ascending
	07/10/2010	HH	Ascending
	22/11/2010	HH	Ascending

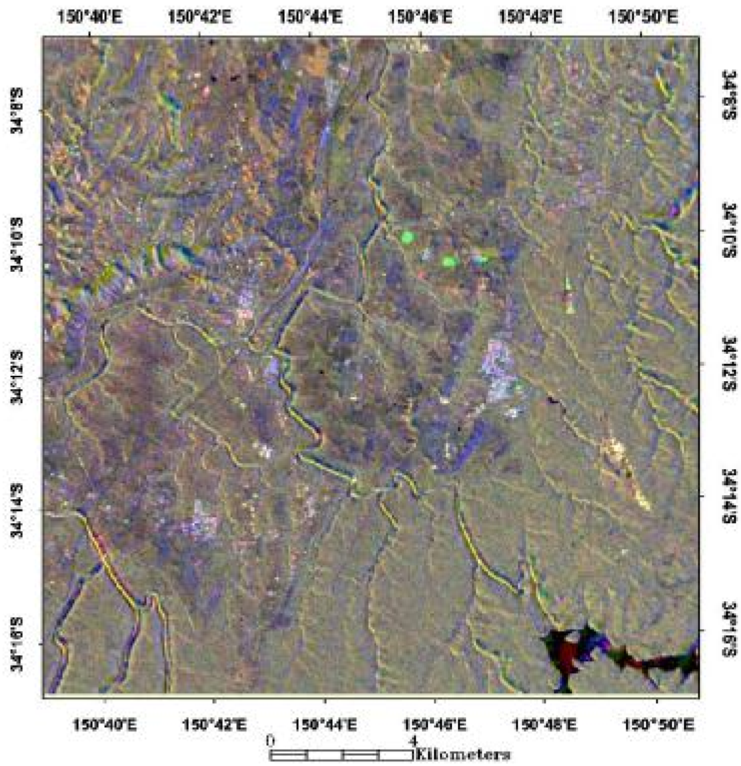


Figure 1:. Color composite of three ENVISAT/ASAR VV polarized images over the study area (acquired on 24/06/2010 (Red), 27/06/2010 (Green) and 25/09/2010 (Blue))

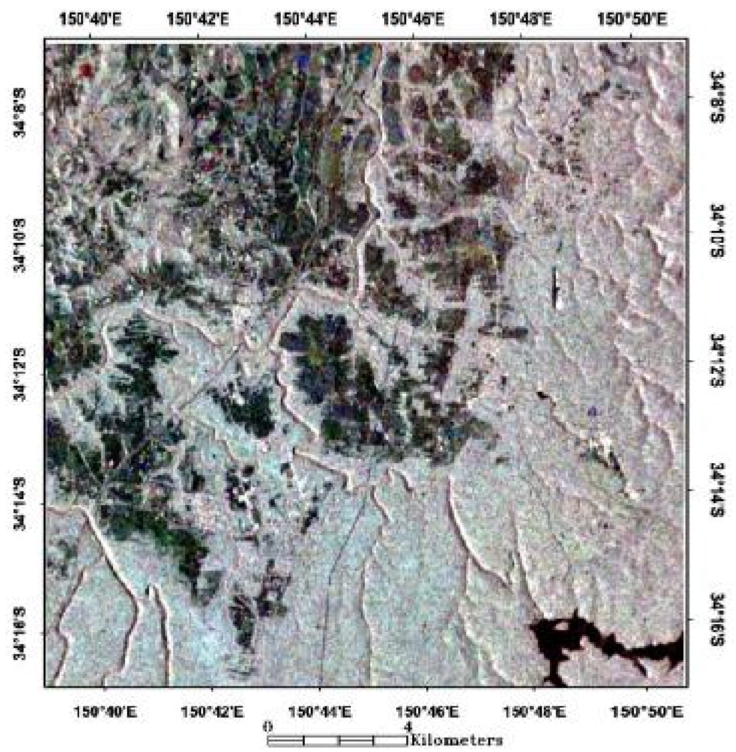
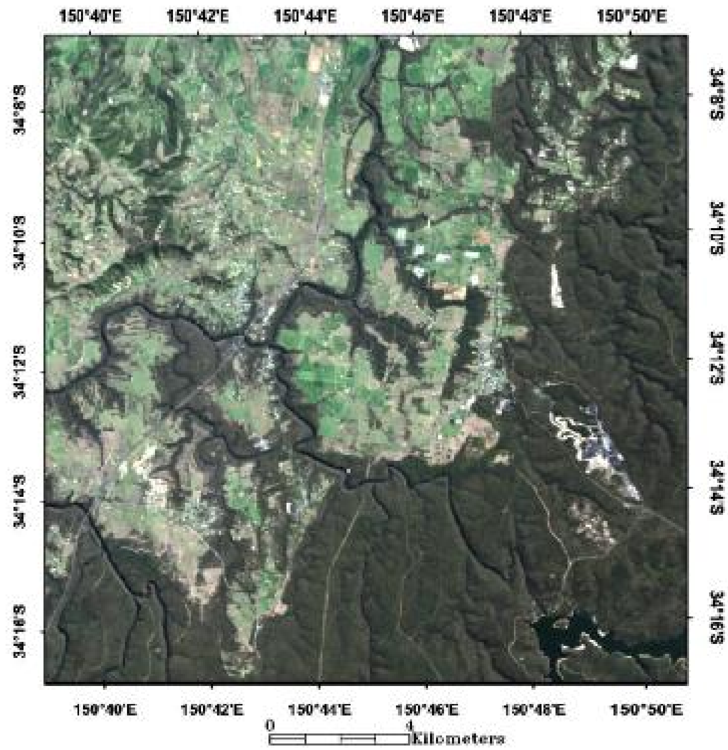


Figure 1:. Color composite of three ALOS/PALSAR HH polarized images over the study area (acquired on 22/05/2010 (Red), 07/07/2010 (Green) and 22/08/2010 (Blue))



Landsat 5 TM image over the study area acquired on 10/09/2010

3. METHODOLOGY

Data preprocessing

Landsat 5 TM data and both kind of SAR images (ENVISAT/ASAR and ALOS/PALSAR) were geo-referenced to the same coordinate system (WGS 84 datum, UTM projection). The DEM was used in the geometric correction process to remove the relief displacements and adjusted SAR backscatter coefficients. All images were then resampled to 15m of pixel size. Speckle noises in SAR images were filtered using the Enhanced Lee filter with 5x5 window size. Pixel's values in the Landsat 5 TM images were converted to reflectance using Equation (1). SAR backscatter values were converted to decibel (dB) using Equation (2)

$$\rho = \frac{\pi . L_\lambda . d^2}{ESUN_\lambda . \cos\theta_s} \qquad (1)$$

where L_λ is the spectral radiance, d is the Earth-Sun distance in astronomical units, $ESUN_\lambda$ is the mean solar exoatmospheric irradiance, and θ_s is the solar zenith angle in degrees.

$$Db = 10 \times \log_{10}(DN)^2 \qquad (2)$$

where DN is SAR magnitude number provided in digital number, Db is magnitude value in decibel.

Derivative data including textural information and Normalized Difference Vegetation Index (NDVI) were also computed and used to create combine datasets. Texture data is valuable data source which can provide additional information on pattern, and spatial arrangement of features on the ground surfaces. In this study, two group of texture were extracted and employed, including first-order and second-order texture measures. Five First Principal Components (PC1) images were generated from six-date ENVISAT/ASAR, six-date ALOS/PALSAR and three Landsat 5 TM images. These PC1 images are used to derive textural information. In order to reduce correlation within datasets only three first-order texture measures, namely Mean, Variance and Data range, and four GLCM texture measures, including Variance, Homogeneity, Entropy and Correlation, were employed. Since there is no preferred direction, the GLCM texture measures were computed as average of texture measures generated for eight different directions of 0^o, 45^o, 90^o, 135^o, 180^o, 225^o, 270^o, 315^o. Textural data were generated from four window sizes, including 5x5, 9x9, 13x13 and 17x17. Since there is no single window size can represent the whole range of texture information in the image (Coburn et al. 2004), all generated texture measures are used simultaneously. Four combined datasets were generated to evaluate the classification performances as presented in the Table 2.

Table 2: Combined datasets for land cover classification in the study area

ID	Datasets	Number of features
1	Six-date PALSAR + six-date ASAR images	12
2	Three-date Landsat 5 TM images	18
3	Three-date Landsat 5 TM + six-date PALSAR + six-date ASAR images	30
4	Three-date Landsat 5 TM + six-date PALSAR + six-date ASAR + + Landsat 5 TM & SAR's textures + three-date NDVI images	173

Classification algorithms

Three non-parametric classifiers were employed for classification processes including ANN based on Multilayer perceptron- Back Propagation (MLP-BP) algorithm, SOM and the SVM.

Artificial Neural Network (ANN)

The MLP-BP model with three layers (input, hidden and output layer) was employed. The number of input neurons is equal to a number of input features, the number of neurons in the output layer is the number of land cover classes to be classified. The number of neuron in the hidden was determined by the sequential testing and validation process using the training data. The sigmoid function was used as the transfer function. The other parameters were set as follows: maximum number of iteration: 1000; learning rate: 0.01-0.1; training momentum: 0.9.

Support Vector Machine (SVM)

The SVM classifier with a Gausian Radical Basis Function (RBF) kernel has been used because of its highly effective and robust in handling of remote sensing data (Kavzoglu and Colkesen 2009, Waske and Benediksson 2007). In order to ensure the best accuracy the optimal value for the penalty parameter C and the width of the kernel function γ were determined by the well known grid search algorithm with five-fold cross-validation techniques.

Self-Organising Map (SOM)

The number of input neuron is equal to the number of input features for each datasets. The output layer of the SOM was a two dimensional array of 15x15 of neurons (total 225 neurons). The neurons in the input layer and output layer are connected by synaptic weights which are randomly assigned within a range of 0 to 1.

Use of multiple classifier system (MCS)

The MCS can be generated in different ways, including combination of different classifiers or combination of the same classifiers with various versions of input training data. The commonly used techniques of "boosting" and "bagging" are based on manipulating input training sample data. In this investigation both methods are implemented. Firstly, the single SVM, ANN and SOM classifiers were carried out to classify the original combined datasets. Secondly, the MCS techniques were applied using bagging and boosting (Adaboost.M1) algorithm and the combination of three classifiers (SVM, ANN and SOM). For comparison purposes, three combination rules were employed to integrate the classification results, namely the Majority Voting (MV), the Bayesian Sum (Sum) and the evidence reasoning based on Dempster – Shafer theory (DS). In the case of boosting, the commonly used Adaboost.M1 algorithm was selected.

Six land cover classes, namely Native Forest (NF), Natural Pastures (NP), Sown Pastures (SP), Urban Areas (UB), Rural Residential (RU) and Water Surfaces (WS) were identified for classification. The UB class includes residential, commercial, industrial, education or research facilities, tourist developments, gaols, cemeteries or abandoned urban area. The training and testing datasets were selected randomly and independently based on visual interpretation with the help of old land use map and higher resolution data available in Google Earth.

4. RESULTS AND DISCUSSION

The overall classification accuracy and Kappa coefficients for the SVM, ANN and SOM classifiers over different datasets is summarised in Table 3.

Table 3: The classification performance of SVM, ANN and SOM algorithms on different combined multisource datasets

Datasets	Overall classification accuracy (%) and Kappa coefficients					
	SVM		ANN		SOM	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	59.26	0.50	59.39	0.50	56.06	0.46
2	79.01	0.74	75.49	0.70	79.97	0.75
3	81.47	0.77	80.99	0.76	80.03	0.75
4	82.78	0.79	81.70	0.77	78.84	0.74

It is clear that for all classifiers the combined datasets (3rd and 4th) gave significant increase in the classification accuracy compared to the single type datasets (1st and 2nd). The integration of textural information and NDVI images slightly enhanced the classification results of the SVM and ANN classifiers by 1.31% and 0.71%, respectively. However this integration reduced the accuracy by 1.19% in the case of the SOM algorithm. The highest classification accuracy 82.78% (Kappa = 0.79) obtained by the SVM classifier on the 4th dataset (Figure 4). It is appeared that the SVM classifier is the most reliable method, it provided the highest classification accuracy for three out of four datasets and gave only slightly lower accuracy of 0.13% compared to the highest accuracy given by ANN classifier. The ANN classifier is second best classifier, which gave higher classification accuracy than the SOM for three (1st, 3rd and 4th) out of four datasets.

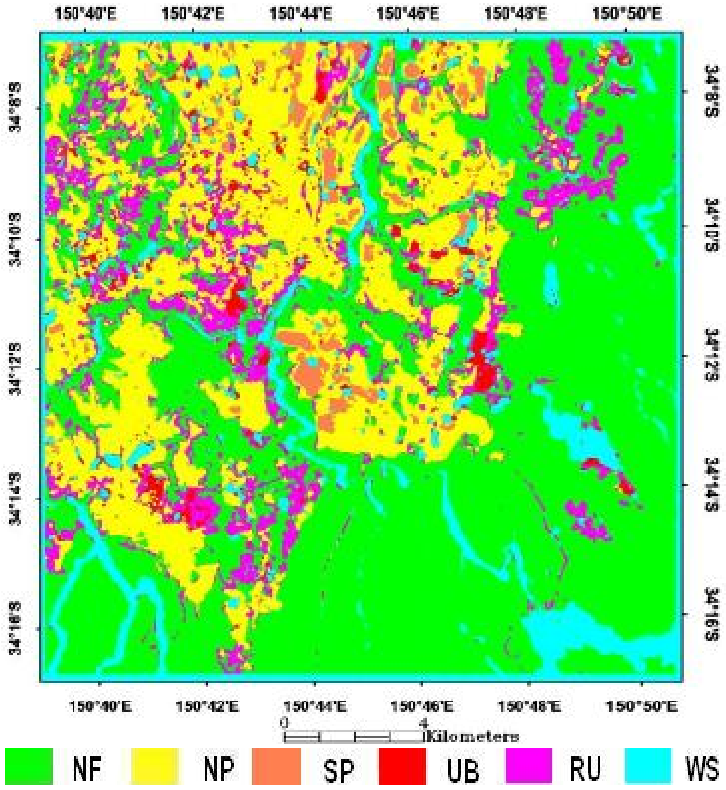


Figure 4: Results of classification using SVM classifier on the 4th dataset.

Classification using Multiple Classifier Systems (MCS)

Bagging and boosting

Results of classification using the bagging and Adaboost.M1 methods with the SVM, ANN and SOM base classifiers are presented in Tables 4 -6 and Figures 5-7.

The bagging and boosting algorithms with ANN and SOM classifiers, in general, gave considerable improvements compared to the performance of the original classifiers. The ANN-Bagging algorithm provided significant increases in overall accuracy (up to 4.8%) for all four datasets. The ANN-Adaboost.M1 gave noticeable improvement in classification accuracy for the 2nd and 3rd datasets, while marginally reduced accuracy for the 1st and 4th datasets. Similarly, the SOM-Bagging algorithm also outperformed the original SOM classifier in all of the datasets. The reason for the success of these MCS techniques is because of the nature of bagging and boosting algorithms can enhance the performance of unstable classifiers such as the ANN or SOM. However, the SOM-Adaboost.M1 produced noticeable decreases in classification performance. The highest classification accuracy was 84.06% for the ANN-Bagging while classifying the 4th dataset. This result is slightly better than the best result of an individual classifier, with 82.78% obtained by the SVM method using the same dataset.

The bagging and boosting (Adaboost.M1) algorithm using the SVM as a base classifier were not superior to the original single SVM classifier in the classification of multisource data over the study. In the case of bagging, the overall classification accuracies decreased for all four datasets. The SVM-Adaboost.M1 algorithm gave some improvements in the 2nd and 3rd datasets, but had reduced classification accuracy for the 1st and 4th datasets. This is not surprising since the SVM classifier is considered to be a stable and accurate algorithm, and consequently the ensemble technique does not help much to enhance the classification performance.

Table 4: Results of classification using single SVM classifier, bagging and Adaboost.M1 techniques based on SVM classifier for different combined datasets

Datasets	Overall classification accuracy (%) and Kappa coefficients					
	SVM		SVM-Bagging		SVM-Adaboost.M1	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	59.26	0.50	59.20	0.50	58.98	0.50
2	79.01	0.74	74.27	0.68	80.93	0.77
3	81.47	0.77	74.72	0.69	82.37	0.78
4	82.78	0.79	80.38	0.76	80.42	0.76

Table 5: Results of classification using single ANN classifier, bagging and Adaboost.M1 techniques based on ANN classifier for different combined datasets

Datasets	Overall classification accuracy (%) and Kappa coefficients					
	ANN		ANN-Bagging		ANN-Adaboost.M1	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	59.39	0.50	64.19	0.55	59.04	0.49
2	75.49	0.70	77.09	0.7168	78.56	0.74
3	80.99	0.76	81.02	0.7662	80.35	0.76
4	81.70	0.77	84.06	0.80	82.91	0.79

Table 6: Results of classification using single SOM classifier, bagging and Adaboost.M1 techniques based on SOM classifier for different combined datasets

Datasets	Overall classification accuracy (%) and Kappa coefficients					
	SOM		SOM-Bagging		SOM-Adaboost.M1	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	56.06	0.46	58.46	0.49	52.90	0.419
2	79.97	0.75	80.86	0.76	78.02	0.7301
3	80.03	0.75	82.34	0.78	79.52	0.7485
4	78.84	0.74	79.30	0.75	77.76	0.7267

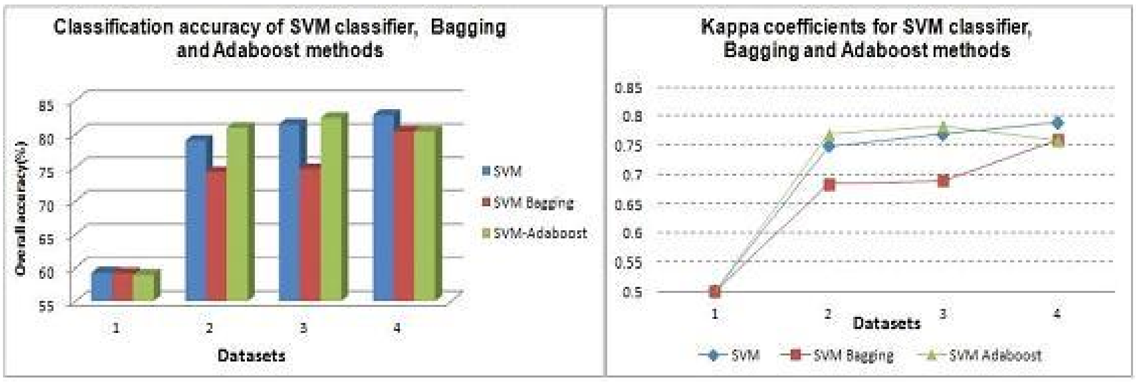


Figure 5: Comparison of SVM classification with SVM-Bagging and SVM-Adaboost.M1 methods

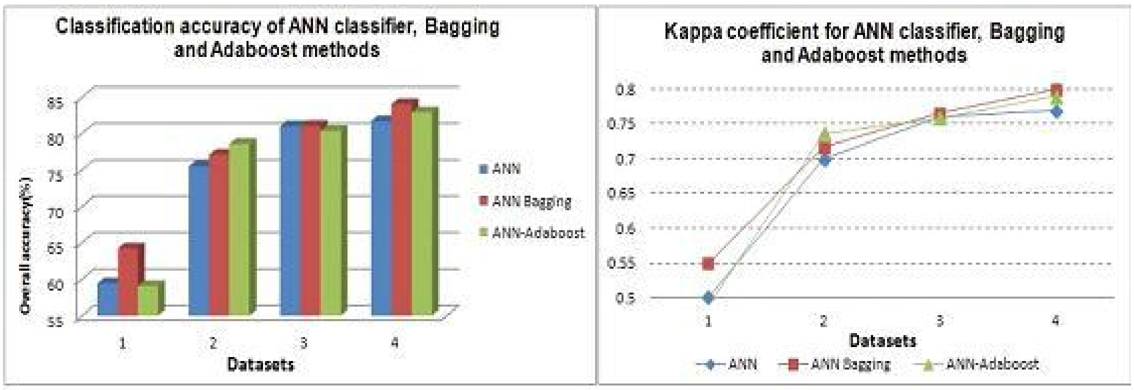


Figure 6: Comparison of ANN classification with ANN-Bagging and ANN-Adaboost.M1 methods

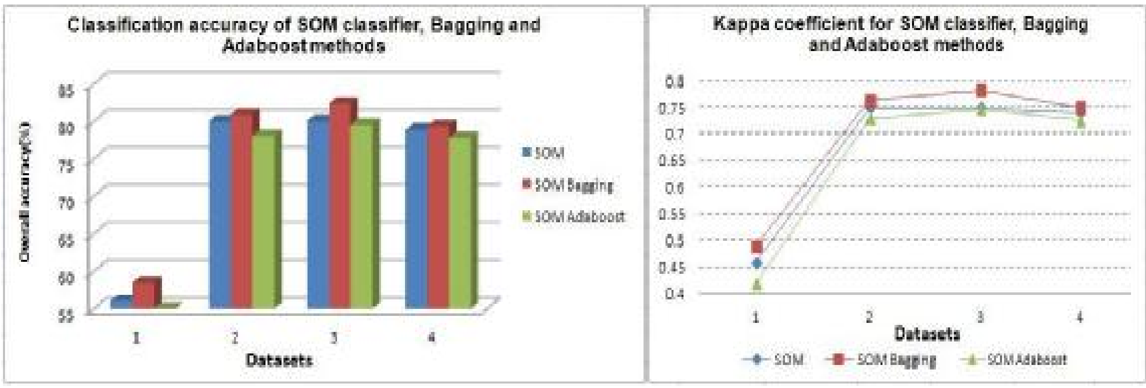


Figure 7: Comparison of SOM classification with SOM-Bagging and SOM-Adaboost.M1 methods

MCS using Majority Voting, Sum and Dempster-Shafer theory combination rules

Results of combination of three classifiers for different datasets are presented in Table 7 and Figure 8 along with results of best individual classifier (BIC) for each dataset. It is clear that the MCS approach outperformed the BICs in most cases. The MCS using Majority Voting and Sum rules produced significantly higher classification accuracy then the best individual classifiers in all cases, while the MCS based on the Dempster-Shafer rule gave better performance in three out of four datasets. The performance of the three decision rules were rather comparable, however the MV and SM rules gave marginally better accuracy than the DS method. It is also worth mentioning that the results of the MCS using a combination of three different classifiers provided consistently higher classification accuracy than the bagging and boosting methods. This superior classification performance can be explained by the higher level of diversity of the different classifiers. The only exception is for the 1st dataset where the ANN-Bagging approach gave the highest classification accuracy of 64.19%, while the MCS using the Dempster-Shafer combination rule gave rather low accuracy of 57.76%.

Table 7: Comparison between the best classification results obtained by individual classifiers and MCS based on the decision rules of Majority Voting, Sum and Dempster-Shafer theory

Datasets	Overall classification accuracy (%) and Kappa coefficients							
	BIC		MV		Sum		DS	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	59.39	0.50	60.32	0.51	59.49	0.50	57.76	0.48
2	79.97	0.75	81.41	0.77	81.06	0.77	80.51	0.76
3	81.47	0.77	83.17	0.79	83.52	0.80	82.98	0.79
4	82.78	0.79	84.77	0.81	84.48	0.81	84.54	0.81

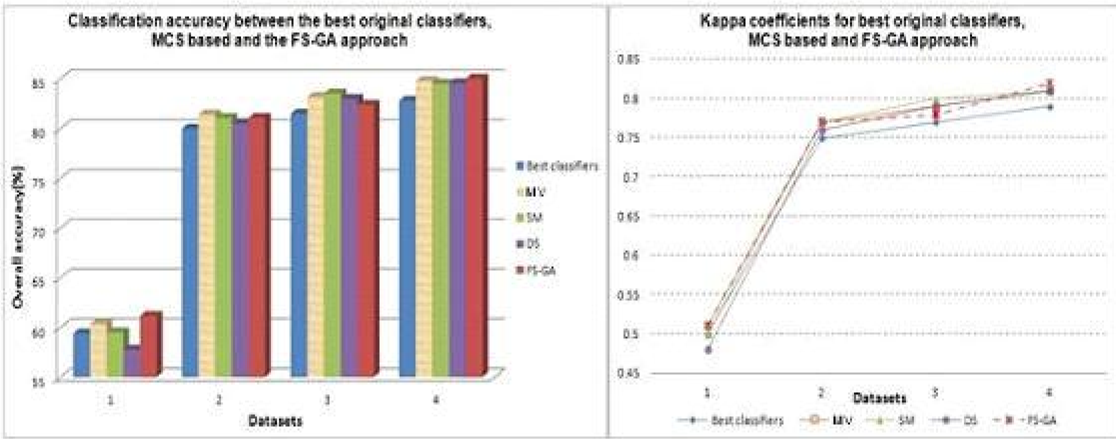


Figure 8: Comparison between the best classification results obtained by original classifiers, MCS based and FS-GA approaches

5. CONCLUSIONS

In this study, several MCS (classifier ensemble) techniques have been used to classify different combined datasets of multi-date ENVISAT/ASAR, ALOS/PALSAR and Landsat 5 TM images. Results illustrated that the MCS technique using combination of different classifiers (SVM, ANN and SOM) is very powerful. This kind of MCS technique outperformed all other methods including individual classifiers, bagging and boosting algorithms. This method gave significant increases in classification accuracy no matter which MV, Sum or DS combination rules were applied. The highest accuracy of 84.77% (Kappa= 0.81) was achieved by using MCS with the MV rule for classifying the largest datasets (which includes 173 features). The performances of three combination rules MV, Sum and DS are relatively comparable. The bagging algorithm worked well with the unstable classifiers, including ANN and SOM. The ANN-Bagging and SOM-Bagging gave a stable increase in the classification for all of four datasets compared to the single base classifier. On the contrary, this algorithm decreased the performance of the SVM classifiers in all cases. The Adaboost-M1 algorithms provide higher classification accuracy than the single classifier only in a few cases. Finally, the investigation also highlighted the usefulness of multisource remote sensing data. In all cases, the combined datasets gave higher classification accuracy than the single-type datasets.

REFERENCES:

Benediktsson, J.A., Chanussot, J., and Fauvel, M., 2007. Multiple classifier systems in remote sensing: from basics to recent developments. In: M. Haindl, J. Kittler and F. Roli, eds.Multiple Classifier Systems. Heidelberg, Germany: Springer, pp. 501–512.

Chu, H.T., and Ge, L., 2010a. Synergistic use of multi-temporal ALOS/PALSAR with SPOT multispectral satellite imagery for land cover mapping in the Ho Chi Minh city area, Vietnam. Proceeding of the International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu HI, 25-30 July, pp. 1465-1468.

Chust, G., Ducrot, D., and Pretus, J. LL., 2004. Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape. International Journal of Remote Sensing 25 (17), pp. 3513–3528.

Coburn, C. A., and Roberts, A. C., 2004. A multiscale texture analysis procedure for improved forest stand classification, International Journal of Remote Sensing 25 (20), pp. 4287–4308.

Dixon, B., and Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both. International Journal of Remote Sensing 29 (4), pp. 1185–1206.

Erasmi, S., and Twele, A., 2009. Regional land over mapping in the humid tropics using combined optical and SAR satellite data – a case study from Central Sulawesi, Indonesia. International Journal of Remote Sensing 30 (10), pp. 2465-2478.

Kavzoglu, T., and Mather, P. M., 2003. The use of backpropagating artificial neural network in land cover classification. International journal of Remote sensing 24 (23), pp. 4907-4938.

Kavzoglu, T., and Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification, International Journal of Applied Earth Observation & Geoinformation 11 (5), pp. 352–359.

Lehmann, E., Caccetta, P., Zhou, Z. –S., Mitchell, A., Tapley, I., Milne, A., Held, A., Lowell, K., and McNeill, S., 2011. Forest Discrimination Analysis of Combined Landsat and ALOS-PALSAR Data. Proceeding of the 34th International Symposium on Remote Sensing of Environment, Sydney, Australia, 10-15 April.

Polikar, R., 2006. Ensemble Based Systems in Decision Making, IEE Circuits and Systems Magazine 6 (3), pp. 21-45.

Salah, M., Trinder, J.C., Shaker, A., Hamed, M., and Elsagheer, A., 2010. Integrating multiple classifiers with fuzzy majority voting for improved land cover classification. ISPRS Int Arch Photogrammetric Engineering and Remote Sensing & SIS 39 (3), Part A, pp.7-12.

Salah, M., Trinder, J. and Shaker, A., 2009. Evaluation of the self-organizing map classifier for building detection from lidda data and multispectral aerial images. Journal of Spatial Science, 54, pp. 15-34.

Soria-Ruiz, J., Fernadez-Ordonez, Y., and Woodhouse, I. H., 2010. Land-cover lassification using radar and optical images: a case study in Central Mexico. International Journal of Remote Sensing 31 (12), pp. 3291–3305.

Tzeng, Y. C., Chiu, S. H., and Chen, K. S., 2006. Improvement of Remote Sensing Image Classification Accuracy by using a Multiple Classifiers System with Modified Bagging and Boosting Algorithms. Proceeding of the International Geoscience and Remote Sensing Symposium (IGARSS), Denver, Colorado, USA, 31 July- 04 August, pp. 2758-2761.

Waske, B., and J. A. Benediktsson, J. A., 2007. Fusion of Support Vector Machines for classification of multisensor data, IEEE Transactions on Geosciences & Remote Sensing, 45 (12), pp. 3858-3866.

Waske, B. & Braun, M. (2009): Classifier ensembles for land cover mapping with multi-spectral SAR imagery. International Journal of Photogrammetry and Remote Sensing 64 (5), pp. 450-457.