Random Forest Algorithm with derived Geographical Layers for Improved Classification of Remote Sensing Data

Uttam Kumar[1, 2], Anindita Dasgupta[3], Chiranjit Mukhopadhyay[1], and T.V. Ramachandra[2, 3, 4, *], *Senior Member, IEEE*
[1]Department of Management Studies, [2]Centre for Sustainable Technologies, [3]Centre for Ecological Sciences,
[4]Centre for *infrastructure*, Sustainable Transport and Urban Planning, Indian Institute of Science, Bangalore – 560012, India.
*Corresponding author: cestvr@ces.iisc.ernet.in

Abstract— **Effective conservation and management of natural resources requires up-to-date information of the land cover (LC) types and their dynamics. The LC dynamics are being captured using multi-resolution remote sensing (RS) data with appropriate classification strategies. RS data with important environmental layers (either remotely acquired or derived from ground measurements) would however be more effective in addressing LC dynamics and associated changes. These ancillary layers provide additional information for delineating LC classes' decision boundaries compared to the conventional classification techniques. This communication ascertains the possibility of improved classification accuracy of RS data with ancillary and derived geographical layers such as vegetation index, temperature, digital elevation model (DEM), aspect, slope and texture. This has been implemented in three terrains of varying topography. The study would help in the selection of appropriate ancillary data depending on the terrain for better classified information.**

Keywords- remote sensing; random forest; classification; land cover; Bangalore; Western Ghats; Himalaya

I. INTRODUCTION

Classification of remote sensing (RS) data accurately is a prerequisite for many environmental and socio-economic applications such as urban change detection, urban heat islands, etc. [1]. Satisfactory classification of RS data depends on several factors including (a) the characteristics of study area, (b) availability of suitable RS data, (c) ancillary and ground reference data, (d) proper use of variables and classification algorithms, (e) user's experience with reference to the application and (f) time constraints [2]. Furthermore, diverse landscapes and terrain types have a mixture of both homogeneous and heterogeneous land cover (LC) classes and require supplemental environmental or geographical layers for improved classification accuracies. Increased spectral variation is common with high degree of spectral heterogeneity in complex landscapes [3]. For example, urban landscapes are composed of features having a complex mix of buildings, roads, flyovers, pavements, trees and lakes which are sometimes smaller than the medium spatial resolution sensors [4]. This creates mixed pixels, a common problem prevalent in residential areas where buildings, trees, lawns, concrete, and asphalt all occur within a pixel, often responsible for low classification accuracy. In landscapes with mountains and dense forests, problems arise due to changes in elevation, topographic differences and often shades (shadows) produced by hillocks and long trees due to altitudinal variations, which is a major challenge for selection of suitable image processing approach over a large area.

On the other hand, availability of fine spatial resolution data such as IKONOS Multispectral (MS) and Panchromatic (PAN) provide vast opportunities in urban studies. A major advantage is the reduction of mixed pixel problem and finer extraction of detailed information of urban entities compared to medium spatial resolution data. However, high spatial resolution data are expensive and requires more time for analysis than medium spatial resolution [2]. Moreover, it often leads to high rate of spectral confusion due to spectral variations present within the LC class and poor classification performance due to limited number of spectral bands [5]. In practice, data acquired from medium spatial resolution sensors such as Landsat TM/ETM+ or IRS LISS-III, being readily available for multiple dates, are commonly used for most landscape analysis (urban and forested terrain at a regional scale). Reducing spectral variation within the same LC class and increasing the separability of different LC types are the keys for improved LC classification [3]. In this regard, different approaches such as sub-pixel classification [6], multi-sensor data integration [7], full spectral image classification [8], expert classification, etc. have been used. Traditional per-pixel spectral classification is based only on spectral signatures, but does not make use of rich spatial information inherent in the data [5]. Therefore, deriving information from RS data with ancillary information (acquired or derived environmental layers) would considerably improve classification accuracy.

Earlier, X. Na et al., (2010) [9] used 103 geographical layers to show improvement in LC mapping using Landsat TM bands 1 to 5 and 7, NDVI, EVI, first principal component (PC1) of the six Landsat TM bands as additional predictors, image texture measures (variance, homogeneity, contrast, dissimilarity and entropy) with window size of 3 x 3 pixels and 11 x 11 pixels, DEM, slope, and soil type. Na. Xiaodong et al., (2009) [10] integrated TM data with NDVI, EVI, PC1, slope, soil types, and five texture measures (variance, homogeneity, contrast, dissimilarity and entropy) for classification of marsh area using Classification Trees and Maximum Likelihood

This project was sponsored by Department of Science and Technology and Ministry of Environment and Forests (MoEF), Government of India, New Delhi, India. We are grateful to Global Land Cover Facility, USA for providing the Landsat imagery. We thank Indian Institute of Science, Bangalore for the financial assistance and infrastructure support.

classifier. This highlights that spectral, textural and terrain data with ancillary derived geographical data improved significantly the LC classification accuracy. A. Fahsi et al., (2000) [11] evaluated the contribution and quantified the effectiveness of DEM in improving LC classification accuracies of the different classes by up to 60% using Landsat TM data over a rugged area in the Atlas Mountains, Morocco. J. A. Recio et al., (2011) [12] used historical land use (LU) and ancillary data as a feature in a geospatial framework for image classification and showed improvement in overall classification accuracy for each class. Masocha and Skidmore (2011) [13] used DEM along with ASTER imagery and geo-referenced point data obtained from field to increase the accuracy of invasive species (Lantana camera) mapping using hybrid classifiers (Neural Network (NN) and SVM classifiers along with GIS expert system). The overall accuracy increased from 71% (kappa 0.61) to 83% (kappa 0.77) with NN and from 64% (kappa 0.52) to 76% (kappa 0.67) with SVM.

G. Xian et al., (2008) [14] quantified multi-temporal urban development characteristics in Las Vegas from Landsat and ASTER Data with ancillary data such as NDVI, slope, aspect and temperature. Lu and Weng (2005) [2] demonstrated urban classification using full spectral information of Landsat ETM+ imagery in Marion County, Indiana. PC's of ETM+ MS bands, texture, temperature and data fusion of MS and PAN were considered to improve classification accuracy. They concluded that data fusion of MS and PAN, with texture and temperature as additional layers are useful but high spatial resolution also increases intra-class spectral variations, decreasing the classification accuracy. Most of the above studies have been based on a single landscape/terrain, sometimes focusing on the comparison of classification techniques, or investigating the role of a few layers on the improvement in classification accuracy, and often using commercial data (such as IKONOS, ASTER) for LC analysis. Hence, there is a need for study that uses free RS data (such as Landsat TM/ETM+) along with other geographical layers for LC classification in different landscape/terrain types using an advanced classification technique that is not complex in its implementation and at the same time does not dependent on the underlying data distribution.

The objective of this work is to investigate the role of ancillary and derived geographical layers such as vegetation indices (NDVI and EVI), elevation and derived layers (slope and aspect), texture (angular second moment, contrast, entropy and variance) and PAN band in addition to original MS bands in improving classification accuracy. The algorithm and classification strategy have been tested in three different terrain types with varying characteristics – Greater Bangalore (highly urbanised terrain), Western Ghats (forested with undulating terrain) and Western Himalaya (rugged terrain with temperate climate) for classifying Landsat ETM+ MS bands.

The paper is organised as follows: section II introduces the data and study area. Section III describes Random Forest classification algorithm. Section IV presents various classification results and discussion followed by conclusions in section V.

II. Data and Study area

A. Data - Survey of India (SOI) Topographical Sheets (of 1:50000 and 1:250000 scales) were used to generate base layers. Field data were collected with pre-calibrated handheld GPS. Landsat ETM+ data were downloaded from Global Land Cover Facility (http://www.landcover.org) and SRTM elevation data were downloaded from http://glcf.umiacs.umd.edu/data/. Google Earth images (http://Earth.google.com) were used with the field data for validating classified outputs. Ancillary layers such as elevation and its derived layers such as slope and aspect along with NDVI, EVI and textures (ASM - angular second moment, contrast, entropy and variance) were used in separate experiments during classification.

B. Study area - Three different terrains were selected to test the performance of the classification strategy as briefed below:

(i) A part of the Greater Bangalore City (Fig. 1) which is a highly urbanised area, comprising dense builtup, parks, roads, streets, flyovers, walk ways, open land, etc. Dense and medium urban areas have high surrounding temperatures compared to vegetation patches, parks and lakes [1]. The undulating terrain in the city ranges from 735 to 970 m with varying textures due to different urban structures.

(ii) A part of Central Western Ghats with gentle undulating hills, rising steeply from a narrow coastal strip bordering the Arabian sea to a plateau at an altitude of 500 m with occasional hills rising above 600 to 860 m (Fig. 2).

(iii) A part of Western Himalaya having temperate climate and rugged terrain with altitude ranging from 295 to 6619 m above mean sea level (Fig. 3).

III. Random Forest (RF)

Classifications were performed using RF technique. RF are ensemble methods using tree-type classifiers $h(x,\Theta_k), k=1,....,$ where the Θ_k are i.i.d. random vectors and **x** is the input pattern. They are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It uses bagging to form an ensemble of classification tree [15]. RF is distinguished from other bagging approaches in that at each splitting node in the underlying classification trees, a random subset of the predictor variables is used as potential variables to define split. In training, it creates multiple Classification and Regression Tree trained on a bootstrapped sample of the original training data, and searches only across randomly selected subset of the input variables to determine a split for each node.

It utilises Gini index of node impurity to determine splits in the predictor variables. For classification, each tree casts a unit vote for the most popular class at input **x**. The output of the classifier is determined by a majority vote of the trees that result in the greatest classification accuracy. It is superior to many tree-based algorithms, because it lacks sensitivity to noise and does not overfit.

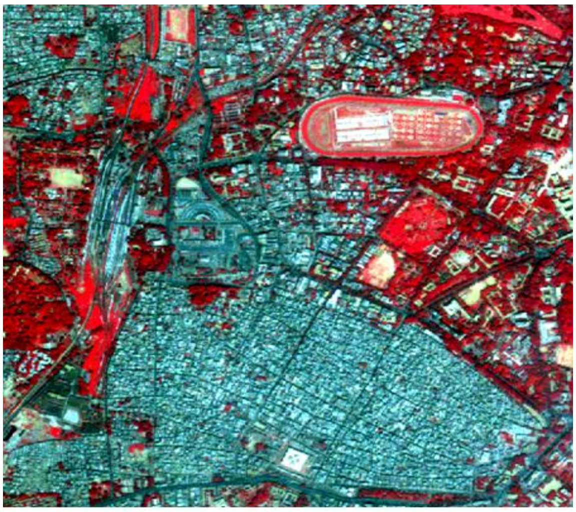

Figure 1. False Colour Composite of the Landsat ETM+ data of a part of Greater Bangalore.

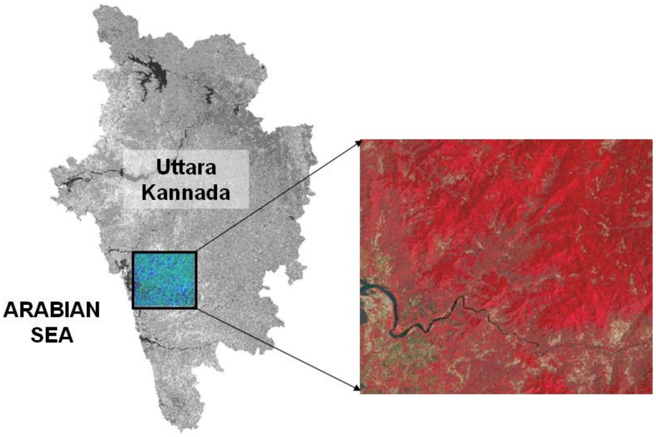

Figure 2. Location of the study area in Central Western Ghats.

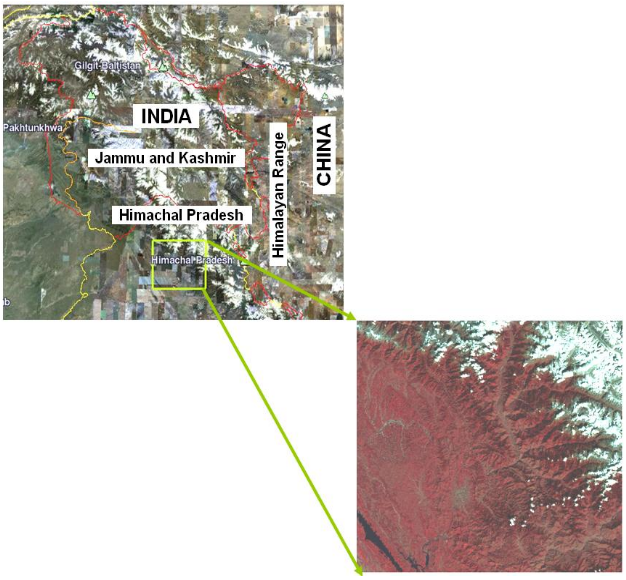

Figure 3. Location of the study area from a part of Western Himalaya.

The trees in RF are not pruned; therefore, the computational complexity is reduced. As a result, RF can handle high dimensional data, using a large number of trees in the ensemble. This combined with the fact that random selection of variables for a split seeks to minimise the correlation between the trees in the ensemble, results in error rates that have been compared to those of Adaboost, at the same time being much lighter in implementation. For more details see [9, 15-16]. Breiman and Cutler (2005) [17] suggests RF "unexcelled in accuracy among current algorithms". RF has also outperformed CART and similar boosting and bagging-based algorithm. In the current work, RF has been implemented using a Linux based random forest package, available in R interface (http://www.r-project.org).

IV. Results and Discussion

A. *Greater Bangalore*

Seven separate classifications summarised in Table I were carried out with different combinations of Landsat ETM+ bands and various geographical layers. Fig. 4 shows classified image with highest overall accuracy (Classification No. 7) and the LC statistics are listed in Table II. The producer's and user's accuracies are not presented due to space constraint. Overall accuracies with kappa statistic are given in Table VI. Outputs obtained from the original spectral bands along with temperature, NDVI, EVI, elevation, slope and aspect (Classification No. 1, 2, 3 and 4) had misclassified many pixels belonging to builtup, water and open area. Water class was over estimated as many paved road (tar or concrete) pixels belonging to builtup were classified as water. Addition of texture, PAN band and texture of PAN significantly improved the classification accuracy of all the classes including urban and water bodies as evident from Table II (highlighted in bold). From accuracy assessment in Table VI it is evident that Classification 5, 6 and 7 have higher accuracies compared to other classifications. Inclusion of temperature increased accuracy whereas addition of vegetation index layers along with elevation, slope and aspect decreased the overall accuracy.

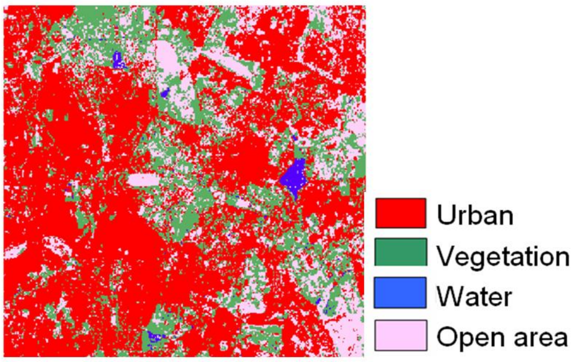

Figure 4. Classified outputs from Landsat ETM+ bands by adding additional geographical layers for Bangalore City.

When both temperature and vegetation index with elevation, slope and aspect were used, the accuracy still decreased. However, inclusion of texture and PAN significantly increased the overall accuracy.

TABLE I.
DETAILS OF DATA AND ANCILLARY LAYERS FOR CLASSIFICATION OF ETM+ DATA FOR A PART OF GREATER BANGALORE

Classification No.	RS data and ancillary geographical layers used	Total number of input layers in the classification
1	ETM+ bands 1, 2, 3, 4, 5 and 7 at 30 m	6
2	ETM+ bands 1, 2, 3, 4, 5, 7 and Temperature	7
3	ETM+ bands 1, 2, 3, 4, 5, 7, NDVI, EVI, elevation, slope and aspect	11
4	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, NDVI, EVI, elevation, slope and aspect	12
5	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, NDVI, EVI, elevation, slope and aspect, texture (ASM, contrast, entropy, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7	108
6	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, NDVI, EVI, elevation, slope and aspect, texture (ASM, contrast, entropy, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7	109
7	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, NDVI, EVI, elevation, slope and aspect, ETM+ PAN, texture (ASM, contrast, entropy, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7, and ETM+ PAN	125

TABLE II.
AREA STATISTICS FOR GREATER BANGAORE CITY

Class → Classification Area ↓		Urban	Vegetation	Water	Open area
1	ha	4543	1912	600	1999
	%	50.17	21.12	6.63	22.08
2	ha	4014	1826	561	2653
	%	44.34	20.17	6.19	29.30
3	ha	4245	1884	471	2446
	%	46.93	20.83	5.21	27.03
4	ha	3450	1854	555	3189
	%	38.14	20.49	6.13	35.24
5	ha	**5263**	**1906**	**94**	**1755**
	%	**58.36**	**21.13**	**1.04**	**19.46**
6	ha	**5226**	**1888**	**87**	**1879**
	%	**57.96**	**20.94**	**0.96**	**20.15**
7	ha	**5164**	**1974**	**66**	**1813**
	%	**57.27**	**21.89**	**0.73**	**20.11**
Total = 9054.62 ha (100%)					

There was 7.6% increase in accuracy by adding temperature, NDVI, EVI, elevation, slope, aspect, PAN along with texture measures (Classification No. 7), which proved to be useful for medium spatial resolution data such as ETM+ while discriminating different classes in an urban environment.

B. Central Western Ghats

Seven classifications were carried out with different combinations of Landsat ETM+ and geographical layers into agriculture, builtup, forest, plantation, wasteland and water bodies that are the six major categories in the forested and mountainous terrain of Uttara Kannada district in Central Western Ghats (Fig. 2). Landsat ETM+ PAN band was resampled to 15 m. The total number of geographical layers was 65 (Table III). Figure 5 shows the classified image with highest classification accuracy (Classificaiton No. 2) and the LC statistics are listed in Table IV. The accuracy assessment table (Table VI) showed that temperature plays a major role in classification in a forested area with highest classification accuracy (88.26%, kappa=0.8643), followed by Classification No. 4 (85.87%, kappa=0.8326). Inclusion of NDVI in spectral bands classification produced very low accuracy, so was removed from further analysis. Addition of elevation, slope and aspect did not improve classification accuracy (Classification No. 3 and 5), and hence were removed from subsequent classifications. Addition of these layers misclassified forest as plantation (> 40% of the area was misclassified as plantation) and wasteland were under estimated. Water bodies could not be detected. Outputs obtained from original spectral bands along with temperature, EVI and PAN (Classification No. 2, 4 and 6) improved the classification results. Texture could not resolve differences between plantation and forest, and plantation was under estimated. EVI increased the classification accuracy by 4.3% (Classification No. 1 and 4) and temperature, EVI and PAN together increased the overall accuracy by 1.6% (in Classification No. 6) compared to the classification of only original spectral bands.

TABLE III.
DETAILS OF DATA AND ANCILLARY LAYERS FOR CLASSIFICATION OF A PART OF CENTRAL WESTERN GHATS

Classification No.	RS data and ancillary geographical layers used	Total number of input layers in the classification
1	ETM+ bands 1, 2, 3, 4, 5 and 7 at 30 m	6
2	ETM+ bands 1, 2, 3, 4, 5, 7 and Temperature	7
3	ETM+ bands 1, 2, 3, 4, 5, 7, elevation	7
4	ETM+ bands 1, 2, 3, 4, 5, 7, EVI	7
5	ETM+ bands 1, 2, 3, 4, 5, 7, elevation, slope and aspect	9
6	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, EVI, PAN	9
7	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, EVI, PAN, texture (contrast, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7, and ETM+ PAN	65

All other layer combinations decreased the accuracy. Overall, the highest classification accuracy improved by 6.7% with temperature as an additional layer.

C. Western Himalaya

Nine classifications were carried out with different combinations of Landsat ETM+ bands and geographical layers (Table V) into four categories - vegetation, water, snow and others (settlement, rock, barren) in the rugged terrain of Western Himalaya (Fig. 3). Landsat ETM+ PAN band was resampled to 15 m.

The total number of geographical layers was 104 (Table V, Classification No. 8). LC statistics, producer's and user's accuracies are not shown here due to space constraint. Overall accuracies and kappa are given in Table VI. Classification of 6 spectral bands and addition of EVI (Classification No. 1 and 4) over estimated the "others" category. In Classification No. 8 (ETM+ bands 1, 2, 3, 4, 5, 7 with temperature, EVI and texture) and 9 (ETM+ bands 1, 2, 3, 4, 5, 7 and texture), others category was not identified. In both these cases, vegetation is over estimated because of misclassification of others category.

TABLE IV.
AREA STATISTICS FOR FOR A PART OF CENTRAL WESTERN GHATS

Class → Area ↓		Agriculture	Built up	Forest	Plantation	Waste land	Water
1	ha	8721	730	37672	4198	886	651
	%	16.5	1.38	71.26	7.94	1.68	1.23
2	ha	8179	854	36978	5195	929	724
	%	15.47	1.61	69.96	9.83	1.76	1.37
3	ha	9315	889	20812	21360	482	-
	%	17.62	1.68	39.37	40.41	0.91	-
4	ha	8973	709	37595	4059	898	624
	%	16.98	1.34	71.12	7.68	1.70	1.18
5	ha	9738	852	19589	22217	463	-
	%	18.42	1.61	37.06	42.03	0.88	-
6	ha	9456	893	36492	4746	604	666
	%	17.89	1.69	69.04	8.98	1.14	1.26
7	ha	8558	8.50	42689	276	823	504
	%	16.19	0.02	80.76	0.52	1.56	0.95
Total = 52858.47 (100%)							

TABLE V.
DETAILS OF DATA AND ANCILLARY LAYERS FOR WESTERN HIMALAYA

Classification No.	RS data and ancillary geographical layers used	Total number of input layers in the classification
1	ETM+ bands 1, 2, 3, 4, 5 and 7 at 30 m	6
2	ETM+ bands 1, 2, 3, 4, 5, 7 and Temperature	7
3	ETM+ bands 1, 2, 3, 4, 5, 7, elevation	7
4	ETM+ bands 1, 2, 3, 4, 5, 7, EVI	7
5	ETM+ bands 1, 2, 3, 4, 5, 7, elevation, slope and aspect	9
6	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, EVI, elevation, slope and aspect	11
7	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, EVI, elevation, slope, aspect and ETM+ PAN	12
8	ETM+ bands 1, 2, 3, 4, 5, 7, Temperature, EVI, texture (ASM, contrast, entropy, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7	104
9	ETM+ bands 1, 2, 3, 4, 5, 7 and texture (ASM, contrast, entropy, variance) at 0, 45, 90 and 135 degrees for ETM+ bands 1, 2, 3, 4, 5, 7	102

The optimum LC classification result with different layers in a rugged terrain is shown in Fig. 6 (Classification No. 7). From accuracy (Table VI), it is evident that addition of each layer in subsequent classifications

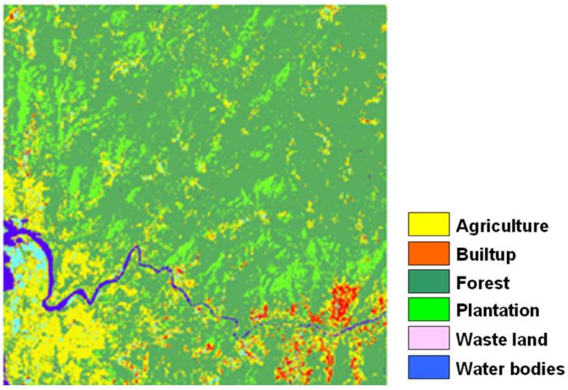

Figure 5. Classified output from Landsat ETM+ bands by adding additional geographical layers for a part of Central Western Ghats.

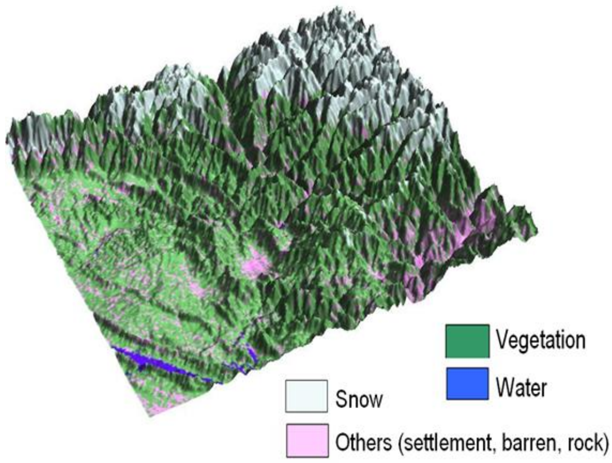

Figure 6. Classified output from Landsat ETM+ bands by adding additional geographical layers for a part of Western Himalaya.

improvised the classification accuracy (Classification No. 2 to 7), compared to only spectral bands (Classification No. 1). Output obtained from original spectral bands along with temperature, EVI, elevation, slope, aspect and PAN showed highest classification accuracy. However, addition of texture did not show any improvement in classification (Classification No. 8 and 9). Addition of temperature layer increased the classification accuracy by 3.76% (Classification No. 1), elevation by 4.18% (Classification No. 3), EVI by 1% (Classification No. 4), elevation, slope and aspect by 5.32% (Classification No. 5) and temperature, EVI, elevation, slope, aspect and PAN together increased the overall accuracy by 10.84% (in Classification No. 7) compared to the classification of only original spectral bands. The analysis revealed that in a rugged terrain with temperate climate, temperature, EVI, elevation, slope, aspect and PAN play major role in improving the classification with highest classification accuracy (Classification No. 7, overall accuracy=89.97, kappa=0.8755). However, texture combinations decreased the accuracy.

TABLE VI.
ACCURACY ASSESSMENT

Classi-fication No.	Study Area					
	Greater Bangalore		Central Western Ghats		Western Himalaya	
	OA*	$\widehat{k}$	OA*	$\widehat{k}$	OA*	$\widehat{k}$
1	75.50	0.7309	81.56	0.7856	79.13	0.7789
2	77.94	0.7548	**88.26**	**0.8643**	82.98	0.7999
3	73.12	0.7101	61.59	0.4931	83.31	0.8122
4	71.43	0.6811	**85.87**	**0.8326**	80.18	0.7865
5	**81.84**	**0.7978**	61.69	0.4874	84.45	0.8222
6	**82.89**	**0.8077**	**83.16**	**0.8014**	87.23	0.8511
7	**83.15**	**0.8125**	77.64	0.7552	**89.97**	**0.8755**
8	-		-		78.91	0.7581
9	-		-		77.19	0.7441
***OA-Overall accuracy, $\widehat{k}$-Kappa**						

In this work, the scope of derived and ancillary layers were assessed for their performance in improving classification accuracy in three diverse terrains. The results provided new insights to the likelihood of improved performance of LC classification by use of supplemental layers related to the region along with the RS data. Although it is difficult to identify suitable texture which is dependent on image band and window size for the specific study, appropriate texture measures reduce the spectral variation within the same LC and also improves the spectral separability among different LC classes. Hilly regions are difficult to classify using RS data due to complex surface features. In this work, temperature, EVI, elevation, slope, aspect and PAN played a major role in increasing the classification accuracy to 89.97% (improvement by 10.84%), compared to the classification of only original spectral bands in a rugged terrain with high altitudinal variations. This accuracy is higher than use of fractral dimension data and original ETM+ data in a Chinese subtropical hilly region (accuracy of 80.69%) by Zhu et al., (2011) [18]. However, in addition to the use of ancillary layers such as textural images, selection of different seasonal data with suitable classification algorithms is also needed to improve classification performance [5].

V. CONCLUSIONS

This work has shown that use of spatial information along with ancillary and derived geographical layers is an effective way to improve classification performance, which was demonstrated through implementation in three different terrains. In a highly urbanised area with less vegetation cover and highly contrasting features, inclusion of temperature, NDVI, EVI, elevation, slope, aspect, PAN and texture significantly increased the overall accuracy by 7.6%. In a forested landscape with moderate elevation, temperature was the only factor that increased the LC classification accuracy by 6.7%. In a rugged terrain with temperate climate, temperature, EVI, elevation, slope, aspect and PAN significantly improved the classification accuracy by 10.84% compared to the classification of only original spectral bands.

REFERENCES

[1] T. V. Ramachandra and U. Kumar, "Land Surface Temperature with Land Cover Dynamics: Multi-Resolution, Spatio-Temporal Data Analysis of Greater Bangalore," *Int. J. Geoinformatics*, vol. 5, no. 3, pp. 43-53, 2009.

[2] D. Lu and Q. Weng, "Urban Classification Using Full Spectral Information of Landsat ETM+ Imagery in Marion County, Indiana," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 11, pp. 1275-1284, 2005.

[3] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performances," *Int. J. Remote Sensing*, vol. 28, no. 5, pp. 823-870, 2007.

[4] J. R. Jensen, Remote Sensing of the Environment: An Earth Resource Perspective, Prentice Hall, Upper Saddle River, New Jersey, 2000.

[5] D. Lu, S. Hetrick and E. Moran, "Land Cover Classification in a Complex Urban-Rural landscape with Quickbird Imagery," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 10, pp. 1159-1168, 2010.

[6] T. Rashed, "Remote sensing of within class change in urban neighbourhood structures," *Computers, Environ. & Urban Sys.*, vol. 32, pp. 343-354, 2008.

[7] B. N. Haack, E. K. Solomon, M. A. Bechdol and N. D. Herold, "Radar and optical data comparison/integration for urban delineation: a case study," *Photogramm. Eng. Remote Sens.*, vol. 68, pp.1289-1296, 2002.

[8] J. Stuckens, P. R. Coppin and M. E. Bauer, "Integrating contextual information with per-pixel classification for improved land cover classification," *Remote Sens. Environ.*, vol. 71, pp. 282-296, 2000.

[9] X. Na, S. Zhang, X. Li, H. Yu and C. Liu, "Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic mapper Imagery and Ancillary Geographic Data," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 7, pp. 833-840, 2010.

[10] Na. Xiaodong, Z. Shuqing, Z. Huaiqing, L. Xiaofeng, L. Chunyue, "Integrating TM and Ancillary Geographical Data with Classification Trees for Land Cover Classification of Marsh Area," *Chinese Geographical Sc.*, vol. 19, no. 2, pp. 177-185, 2009.

[11] A. Fahsi, T. Tsegaye, W. Tadesse and T. Coleman, "Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy," *Forest Ecology and Management*, vol. 128(2000), pp. 57-64, 2000.

[12] J. A. Recio, L. A. Hermosilla and A. Fernandez-Sarria, "Historical Land Use as a Feature for Image Classification," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 4, pp. 377-387, 2011.

[13] M. Masocha and A. K. Skidmore, "Integrating conventional classifiers with a GIS expert system to increase the accuracy of invasive species mapping," *Int. J. Applied Earth Observation and Geoinformation*, vol. 13, pp. 487-494, 2011.

[14] G. Xian, M. Crane and C. McMahon, "Quantifying Multi-temporal Urban development Characteristics in Las Vegas from Landsat and ASTER Data," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 4, pp. 473-481, 2008.

[15] L. Breiman, "Random Forests," *Machine Learning*, vol. 40, pp. 5-32, 2001.

[16] L. Breiman and A. Cutler, "Random Forests", 2005. URL: http://www.stat.berkeley.edu/users/breiman/RandomForests/

[17] L. Breiman and A. Cutler, "Breiman and Cutler's random forests for classification and regression," Version 4.5-36, Repository – CRAN, 2010.

[18] J. Zhu, J. Shi, H. Chu, J. Hu, X. Li and W. Li, "Remote Sensing Classification Using Fractal Dimension over a Subtropical Hilly Region," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 1, pp. 65-74, 2011.