

# EXTRACTION OF BUILT-UP AREA BY COMBINING TEXTURAL FEATURES AND SPECTRAL INDICES FROM LANDSAT-8 MULTISPECTRAL IMAGE

V. S. Bramhe<sup>1,\*</sup>, S. K. Ghosh<sup>1</sup>, P. K. Garg<sup>1</sup>

<sup>1</sup>Geomatics Engineering Group, Civil Engineering Department, IIT Roorkee, 247667, India,  
vijendra89singh@gmail.com

Commission V, SS: Infrastructure and Development Planning

**KEY WORDS:** Built-up, Support Vector Machine, Texture, Spectral Indices, GLCM.

## ABSTRACT:

Remote sensing techniques provide efficient and cost-effective approach to monitor the expansion of built-up area, in comparison to other traditional approaches. For extracting built-up class, one of the common approaches is to use spectral and spatial features such as, Normalized Difference Built-up index (NDBI), GLCM texture, Gabor filters etc. However, it is observed that classes such as river soil and fallow land usually mix up with built-up class due to their close spectral similarity. Intermixing of classes have been observed in the classified image when using spectral channels. In this paper, an approach has been proposed which uses urban based spectral indices and textural features to extract built-up areas. Three well known spectral indices i.e. NDBI, Built-up Area Extraction Index (BAEI) and Normalized Difference Bareness Index (NDBai) have been used in this work. Along with spectral indices, local spatial dependency of neighborhood regions is captured using eight GLCM based textural feature, such as, Contrast, Correlation, Energy and Homogeneity etc. for each image band. All textural and spectral indices bands are combined and used for extracting built-up areas using Support Vector Machine (SVM) classifier. Results suggest 4.91% increase in overall accuracy when using texture and spectral indices in comparison with 84.38% overall accuracy achieved when using spectral data only. It is observed that built-up class are more separable in the projected spectral-spatial feature space in comparison to spectral channels. Incorporation of textural features with spectral features reduces the misclassification error and provides results with less salt and pepper noise.

## 1. INTRODUCTION

Urbanization shown a rapid growth in past few decades, affecting many dimensional of today's world. It is essential to have the accurate information about the extent and nature of built up land cover in order to support the sustainable growth of in rural and urban areas. Identification of urban expansion, understanding the urban dynamics along with the associated consequences became an important as well as interesting research domain (Karathanassi, 2000; Benediktsson, 2003; Zhou, 2008; Li, 2014). However, accurate mapping of urban area is a critical task and problem becomes more challenging in developing countries due to lack of resources and infrastructure. Remote sensing technology provides feasible and cost effective way to monitor land covered by built-up structures (Pesaresi, 2011). Extraction of built-up areas is a challenging task due to the similarity of spectral properties of classes such as sand and built up area. Initial studies on classifying rural/urban areas using structural information (edge density map in Gong (1990). Advanced image processing transforms raw satellite data into useful maps by multiple stage of processing according to well-defined algorithms. Using satellite images, changes in built-up regions and its impact on the socio-economic and environmental conditions of any geographic area can be monitored with relative ease.

With the availability of Landsat satellite archive i.e. the largest series of space-borne earth observation data as well as at good spatial resolution provides huge opportunities for urban mapping. Many studies shown the suitability of Landsat data for urban mapping and monitoring (Yuan, 2009; Taubenböck, 2012; Al-Doski, 2013; Sexton, 2013; Bhatti, 2014; Becherer, 2017). Spectral indices play an important role in extraction of built up

area using remotely sensed data. Normalized Difference Built-up Index (NDBI) proposed by Zha (2003) is one of the most commonly used index for built up extraction.

In literature, various spectral indices have been proposed such as Band Ratio for Built-up Area (BRBA), Normalized Built-up Area Index (NBAI), New Built-up Index (NBI), Enhanced Built-Up and Bareness Index (EBBI) etc. Although these techniques do not require finding training samples and building classifier on them but it requires finding threshold. Another important features are texture measures used in a number of studies to highlight the built up area (Chen, 2013; Zhang, 2013; Kuffer, 2016). Pesaresi (2008) presented a novel method to compute built-up presence index (PanTex) using textural characteristics of panchromatic data. This study demonstrate that Pantex minimize the edge effect and also improves the capacity to discriminate the built up from rest of other classes. Bouzekri (2015) proposed a new spectral index and compared it with well-known spectral index BRBA, NBAI, NBI and NDBI for built up extraction and result showed the superiority of proposed index. Xu (2008) proposed an index-based built-up index (IBI) for built up area extraction using satellite imagery and results showed that IBI improves built-up extraction and suppress background noise. Bhatti (2014) explored the capability of NDBI and results showed the superiority of proposed built-up area extraction method (BAEM) as over NDBI. As-syakur (2012) proposed EBBI and authors claimed that this index is capable to separate built-up and bare land effectively. NDBI and BAEI highlights built-up regions in the images whereas NDBai highlights bare areas. SVM is a machine learning algorithm received a lot of attention by remote sensing community in last decade because of its fair classification results. Cao (2009) proposed a semi-automatic approach to map built up area and shown its usefulness over manual error and trial

\* Corresponding author

approach. In this study, SVM model has been trained using pixels of known location for all the features extracted. This study aims to explore the usefulness of selected GLCM texture measure to extract built-up area using Landsat-OLI data. Eight GLCM texture measure (Table 1) have been computed. On the other hand, built-up spectral indices are also incorporated in input image and its usefulness is also evaluated for urban area mapping.

## 2. STUDY AREA AND DATA USED

Due to the availability of Landsat-8 data with free and global coverage since 2013, given a huge opportunity to a wide range of remote sensing applications like vegetation mapping, built-up extraction, fire detection, hydrological modeling etc. The selected study area is Haridwar Tehsil located in Uttarakhand, India. This study use Landsat-8 OLI image acquired on 13 November 2016 and covered area comprises of 1166.35 sq. kms. The satellite data used in this study consists of multispectral bands acquired by Landsat 8 OLI sensor. The image represents a diverse land class scenario with  $1444 \times 1744$  pixels in seven bands ranging from the wavelength of 0.43-2.29 in the spectrum and having a spatial resolution of 30 m. In the last few decades, rapid urbanization has been taken place in this area, which results in increased infrastructural growth and urban expansion. The area comprises heterogeneous land cover types including built-up regions, agricultural area, water, sand and fallow land. The False-Color Composite (FCC) image of the study area is shown in Figure 1.

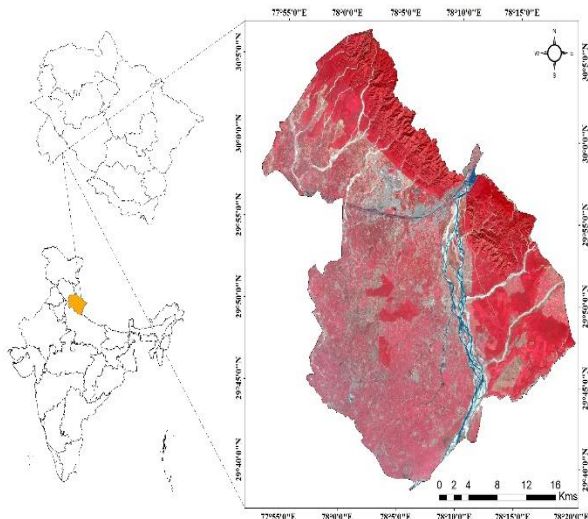


Figure 1. FCC of the study area Haridwar, Uttarakhand, India

In this study, besides using optical multispectral data, The Phased Array Type L-band Synthetic Aperture Radar (PALSAR) Digital Surface Model (DSM) jointly provided by JAXA and the Japan Resources Observation System Organization (JAROS) has also been utilized. This PALSAR DSM is used to generate slope and aspect of the study area.

All the datasets have been spatially synchronizing with each other so that there will be no location shift occur. The projection system has taken as UTM for all the datasets.

## 3. METHODOLOGY

The methodology adopted for mapping of built-up area in this study is shown in Figure 2. Firstly, the Landsat 8 OLI data is utilized for extraction of various GLCM based textural measures and spectral indices (i.e. NDBI, NDBai and BAEI). Textural features at multiple kernel scales are generated and on the basis of sum of mutual information score of individual textural feature, highest information textural features are selected. Generated spectral and textural features are incorporated with each other to form a spectral-spatial feature set. Along with these spectral-spatial features, ancillary data i.e. topographic parameters (slope and aspect) and Land surface temperature (LST) are also incorporated as input feature set. Random Forest (RF) feature selection method is utilized for selection of optimal features from the generated feature set for the mapping of built-up areas.

SVM is a robust machine learning classifier, is used to classify the image. Ground truth data have been collected by field visit of the study region. For accuracy assessment, confusion matrix based statistical measures i.e. overall accuracy, omission/commission error and kappa coefficient have been used. The proposed approach has been implemented using various python libraries (scikit learn, tensorflow, gdal/ogr etc.)

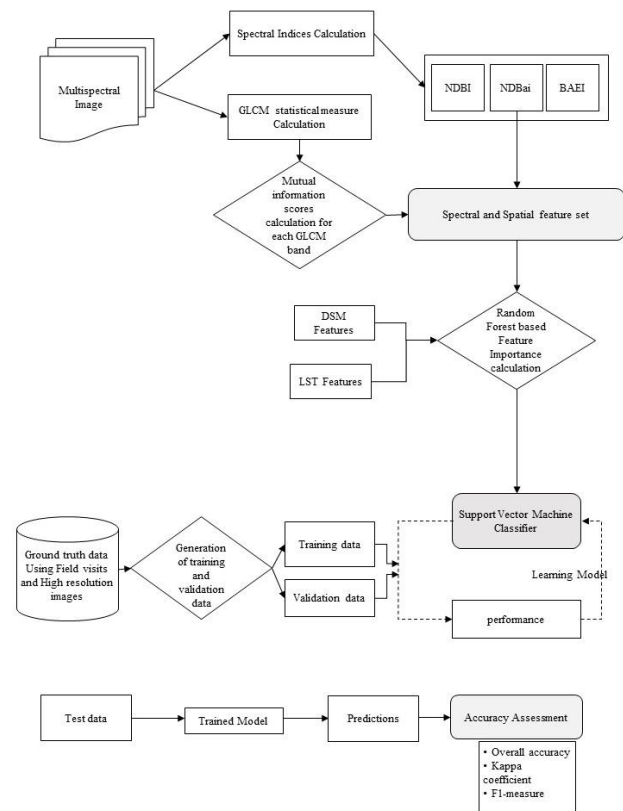


Figure 2. Flowchart of proposed approach for mapping of built-up area

In the proposed approach textural measures, spectral indices are discussed below:

### 3.1 Built-up Spectral Indices

Spectral indices are widely used in various applications of remote sensing due to providing relevant class information with less computational overload. The built-up indices are specific to

highlight built-up class in a satellite image which is also helpful in mapping built-up regions. These indices such as NDBI or NDBai (bareness index) are often useful in built-up extraction task because these will be helpful in minimizing the class confusion. In this study, we have calculated three commonly used spectral indices namely NDBI, BAEI and NDBai. Which are discussed in next subsections.

### 3.1.1 Normalized Difference Built-up index (NDBI)

NDBI is basically proposed for Landsat Thematic Mapper (TM) images by Zha (2003) for creating built-up index image. It utilizes SWIR and NIR bands for calculation as depicted in equation (1)

$$NDBI = \frac{(SWIR - NIR)}{(SWIR + NIR)} \quad (1)$$

### 3.1.2 Built-up Area Extraction Index (BAEI)

BAEI is proposed by Bouzekri (2015) for highlighting built-up areas in Landsat-8 image. Apart from using spectral channels, an arithmetic constant is also introduced to facilitate the extraction as shown in equation (2)

$$BAEI = \frac{Red + L}{Green + SWIR1} \quad (2)$$

L = an arithmetic constant 0.3

### 3.1.3 Normalized Difference Bareness Index (NDBai)

NDBai is first proposed by Hongmei (2005) for extraction of bare land from Landsat 5 TM and Landsat 7 ETM+ images. It makes use of SWIR-1 and Thermal bands as shown in equation (3)

$$NDBai_{1,2} = \frac{SWIR_{1,2} - TIRS_{1,2}}{SWIR_{1,2} + TIRS_{1,2}} \quad (3)$$

Since, Landsat 8 consists of two SWIR and TIRS bands therefore both bands can be taken simultaneously and the mean,  $NDBai_{mean}$  of 1 and 2 can be taken.

## 3.2 GLCM Textural features

Satellite images also contains rich information in terms of context or spatial information, apart from having spectral information encoded in different wavelength channels. The context or neighborhood also play a crucial role in mapping of built-up areas. Since, each class or object pixels shows distinguished characteristics in accordance with the neighboring pixels. Thus, finding spatial information may result in increasing the accuracy of classification to a large extent. In this study, we have calculated GLCM based textural measures for capturing the spatial relationship exists in the image. GLCM measure are widely used statistical quantities because these are calculated using co-occurrence matrix which provides the basis for calculating multiple first or second order statistical quantities such as entropy, homogeneity etc. In Table 1 eight GLCM measure calculated in this study are shown.

S. No.	Feature Extracted	Formula
--------	-------------------	---------

1	Mean ( $\mu$ )	$\mu_i = \sum_{i,j} i(p(i,j))$ $\mu_j = \sum_{i,j} j(p(i,j))$
2	Variance ( $\sigma^2$ )	$\sigma_i^2 = \sum_{i,j} (i - \mu_i)^2 p(i,j)$ $\sigma_j^2 = \sum_{i,j} (j - \mu_j)^2 p(i,j)$
3	Homogeneity	$\sum_{i,j} \frac{p(i,j)}{1 +  i-j ^2}$
4	Contrast	$\sum_{i,j}  i-j ^2 p(i,j)$
5	Dissimilarity	$\sum_{i,j}  i-j  p(i,j)$
6	Entropy	$\sum_{i,j} -p(i,j) \log(p(i,j))$
7	Second Moment	$\sum_{i,j} p(i,j)^2$
8	Correlation	$\sum_{i,j} p(i,j) \left[ \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$

Table 1. GLCM based textural measures

## 4. RESULTS AND ANALYSIS

In this section, experimental setup, obtained results and analysis has been discussed. First, GLCM textural features are generated and selected based on highest sum of mutual information. These features are incorporated with spectral indices, topographic and LST features and provided as input to non-parametric SVM classifier.

### 4.1 Generation and Selection of GLCM textural feature

Various GLCM based textural features as discussed in section 3.2 is calculated for each 7 spectral bands of Landsat OLI data. There are multiple parameters involved in calculation of GLCM matrix such as kernel size, offset direction, quantization level and spatial neighborhood. Due to large number of possible combinations of parameter involved in selection of best textural measures, a huge computational cost may incur. Therefore, some of parameters i.e. spatial neighborhood which defines the spatial relationship between reference pixel and the neighboring pixel is chosen as immediate neighbour and quantization level which assign a discrete gray level value for the calculation of GLCM is taken as 64.

Kernel size is one of the most important parameter which defines the context around the reference pixel (Hall-Beyer, 2017). Kernel size takes the number of neighboring pixels need to be considered around center pixel. For selection of highly informative textural features, 5 different kernel sizes i.e. 3x3, 7x7, 11x11, 15x15 and 21x21 are considered. Table 2 shows sum of mutual information of each textural features for 5 different kernel sizes. Average of 4 offset direction [0°, 45°, 90° and 135°] in a kernel window is considered to calculate all texture features. Average of these directions have been taken to provide directional independence to outcome features. Offset direction is

chosen to be mean of all the directions (i.e. 0°, 45°, 90° and, 135°) hence the output features are directional independent.

Kernel size	3 x 3	7 x 7	11 x 11	15 x 15	21 x 21
Sum of mutual information	13.956252	18.73362	19.57776	19.21531	18.02559

Table 2. Sum of mutual information feature score for 5 different kernel sizes.

Highest information score is obtained by using the kernel size of 11x11, therefore it is selected for the further study for built-up area extraction. The textural map of is shown in in Figure 3.

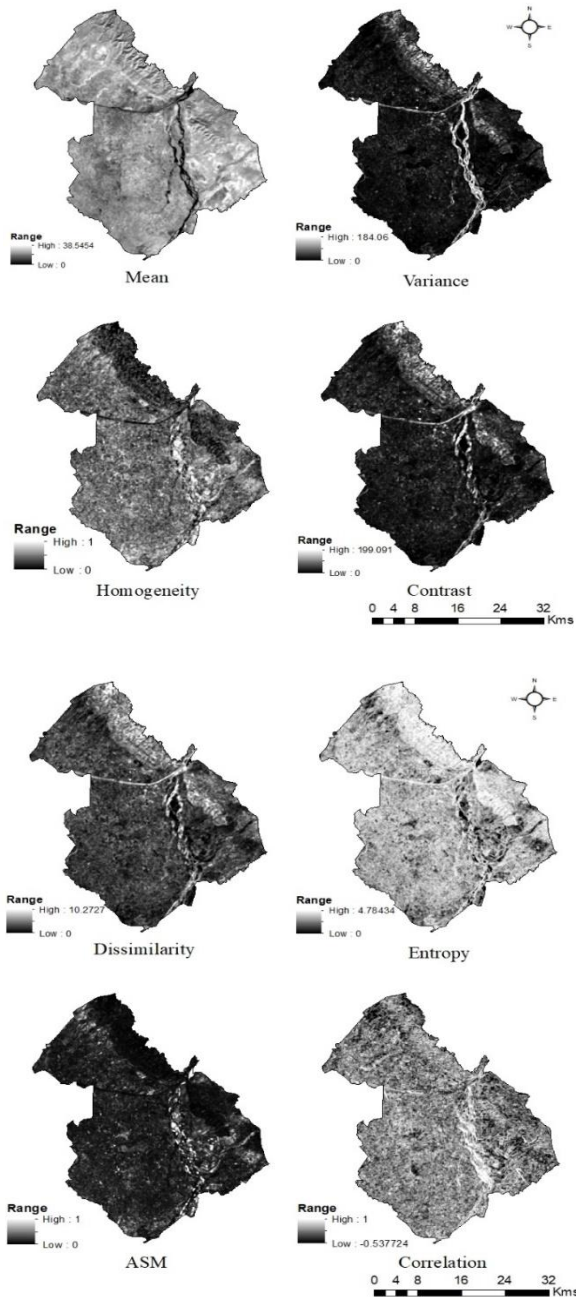


Figure 3. Textural feature maps of various GLCM measures obtained using kernel size of 11 x 11.

#### 4.2 Calculation of Spectral Indices

Spectral Indices utilizes spectral reflectance properties to distinguish classes present in the image. Built-up area indices as discussed in section 3.1 have been generated using Landsat 8 data and used as an input feature to the non-parametric SVM classifier for learning to extract built-up areas. Since, there is a lot of spectral confusion between built-up class and bare soil therefore apart from using two built-up indices, a bareness index is also calculated. The three spectral indices computed are shown in Figure 4.

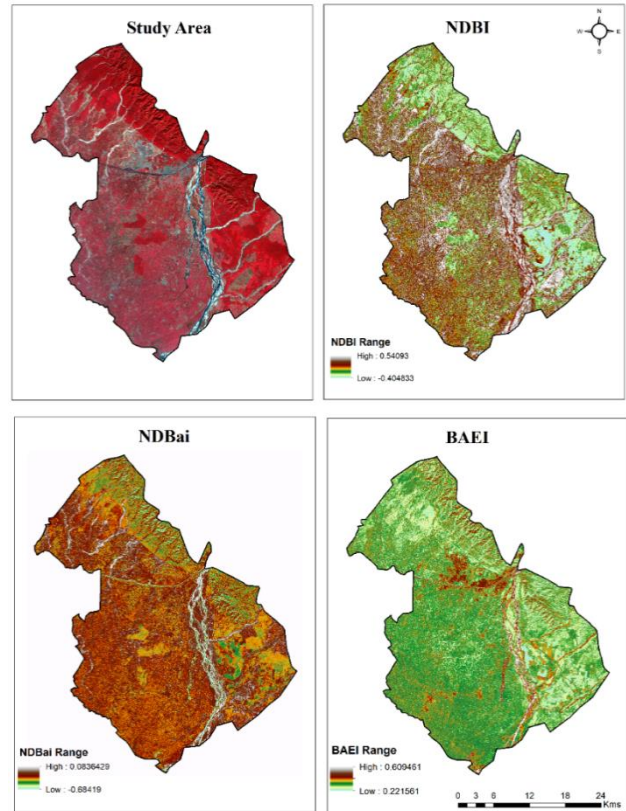


Figure 4. (a) Original OLI image (b) NDBI (c) NDBai (d) BAEI

Visual analysis of extracted spectral indices shows that BAEI gives very good results than NDBI for highlighting built-up areas in the image. In BAEI, dense built-up areas comes up very clearly. Whereas, NDBai highlights sandy areas in the present in the image. Although, calculation of spectral indices are computationally easier, one of the major drawback of utilizing spectral indices directly in remote sensing application is that it is very difficult to find the thresholds to distinguish the classes.

#### 4.3 Generation of Topographic features and Land surface temperature

Apart from, spectral indices and textural features, another set of features which are taken as predictors in SVM classifier are topographic features (slope and aspect) and LST (Figure 5). Thermal bands available in Landsat 8 data can capture energy in wavelength range of 10.60 to 12.51 μm hence can be used for calculating surface temperature. Bare land, semi-bare land and land under development are warmer than other land cover types. Hence LST may contribute to distinguish different land use land

cover classes. However, in this work LST feature did not show much impact to the accuracy of built up mapping

In urban studies topographic. Similarly, topographic parameters slope and aspect also did not improve the accuracy of the classification.

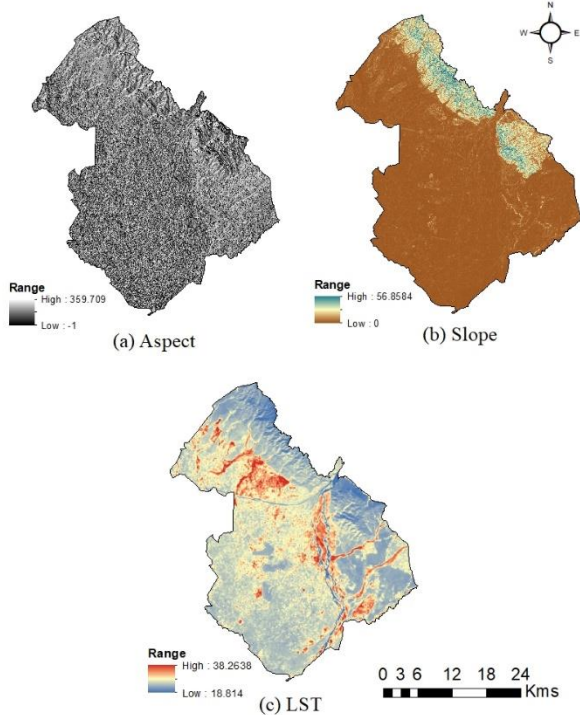


Figure 5. (a) Aspect, (b) Slope (c) represents average Land Surface Temperature (LST) map of the Study Area

#### 4.4 Feature importance of individual feature using Tree based approach

Each feature either textural or spectral, calculated from the original raw spectral dataset depicts some preference in the classification of built-up areas. Hence we have calculated feature importance of each variable calculated and instead of using all the features or random selection of feature, those features are selected which shows higher gain in performance as shown in Figure 6.

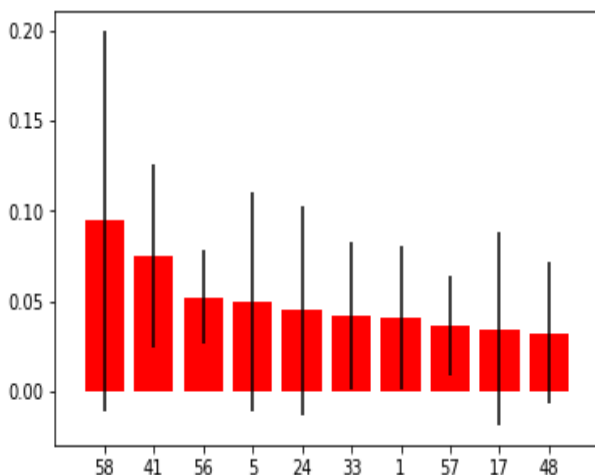


Figure 6. Gini importance scores of selected features

Random forest method is utilized for feature importance calculation. Feature importance is calculated on the basis of Tree based classifier. In the feature set, all generated features are stacked in such a way that Feature No. 1 to 56 are texture features, Feature No. 57 to 59 are BAEI, NDBai and NDBI respectively, while Feature No. 60-62 are Aspect, slope and LST respectively. Thus, a total of 62 features are stacked together. Features and their Gini importance score is shown in table 3.

S. No.	Selected feature	Gini importance score
1.	feature 58	0.094337
2.	feature 41	0.074816
3.	feature 56	0.052127
4.	feature 5	0.049592
5.	feature 2	0.044682
6.	feature 33	0.041848
7.	feature 1	0.041153
8.	feature 57	0.036104
9.	feature 1	0.034417
10.	feature 48	0.032349

Table 3. Best features having highest Gini importance score

#### 4.5 Accuracy assessment using test data and built-up map generation

Here, measures used in assessment of accuracy are discussed. These metrics have been calculated for test data on which prediction has been done using trained machine learning (SVM) model. There are multiple quantitative measures available to compute the effectiveness of trained model. For quantitative analysis, non-overlapping samples or mutually exclusive samples from training and validation set has been taken and overall accuracy, kappa coefficient and f-1 measure are calculated.

In this work, SVM classifier has been implemented for built-up extraction using Landsat-8 OLI imagery. Results shows that overall accuracy is 84.38% when only spectral bands are used as input features whereas kappa coefficient is .6004 (Table 4). However, a rise of 4.91% has been observed when bands are selected by using RF feature selection method and similarly a significant improvement is noticed in kappa value as it reaches to .7556. Another accuracy measure i.e. F1 score gives a value of .8609 when only spectral bands are used and maximum value of .9077 when classification is performed on bands selected by RF selection method (Table 4). The obtained classified map showing built-up and non-built-up areas is shown in Figure 7. It can be seen that when using spectral bands only a lot of salt and pepper noise appears whereas in case of using features selected using Random Forest feature importance misclassification rate reduces.

	Spectral bands only	Bands Selected using RF feature selection
Overall Accuracy	84.38	89.29
Kappa Coefficient	0.6004	0.7556
F1-score	0.8609	0.9077

Table 4. Accuracy assessment on test dataset from spectral features and features selected using Random forest classifier

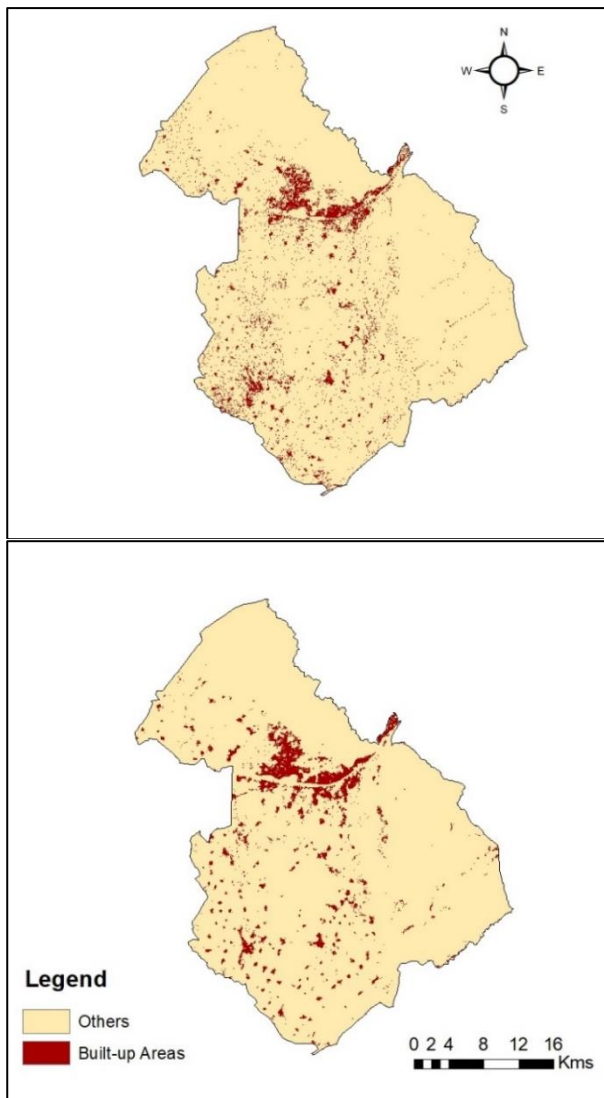


Figure 7. (a) Built-up area map using Landsat OLI spectral bands (b) built-up area map using selected features

## 5 CONCLUSIONS

This study analyzes the impact of GLCM based textural measures and selected built-up indices on built-up extraction using Landsat-8 OLI data. Results indicates that NDBai shows highest importance followed by BAEI. In this work, different kernel sizes have been tested and it is observed that bigger kernel causes smoothing effect on calculated texture measure. Smaller kernels are capable to capture smaller shape boundaries whereas larger kernel diminishes smaller shapes, it is because of considering more number of pixels. Kernel size of 11x11 gives highest information content for built-up extraction. Topographic parameters did not shown much significance for mapping of built-up area because topographic changes of study area is very less. At the same time LST did not shown any contribution because temperature variation is almost ranging near mean value of temperature throughout the study area. Highest overall accuracy has been achieved by using features selected by Random forest method over spectral bands only.

## REFERENCES

- Al-Doski, J., Mansor, S.B., Shafri, H.Z.M., 2013. Support vector machine classification to detect land cover changes in Halabja City, Iraq. In: *BEIAC 2013 - 2013 IEEE Bus. Eng. Ind. Appl. Colloq.*, pp. 353–358.
- As-syakur, A.R., Adnyana, I.W.S., Arthana, I.W., Nuarsa, I.W., 2012. Enhanced built-UP and bareness index (EBBI) for mapping built-UP and bare land in an urban area. *Remote Sens.*, 4(10), pp. 2957–2970.
- Becherer, N., Pecarina, J., Nykl, S., Hopkinson, K., 2017. Improving optimization of convolutional neural networks through parameter fine-tuning. *Neural Comput. Appl.*, pp. 1–11.
- Benediktsson, J.A., Pesaresi, M., Arnason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Trans. Geosci. Remote Sens.*, 41(9), pp. 1940–1949.
- Bhatti, S.S., Tripathi, N.K., 2014. Built-up area extraction using Landsat 8 OLI imagery. *GIScience Remote Sens.*, 51(4), pp. 445–467.
- Bouzekri, S., Lasbet, A.A., Lachehab, A., 2015. A New Spectral Index for Extraction of Built-Up Area Using Landsat-8 Data. *J. Indian Soc. Remote Sens.*, 43(4), pp. 867–873.
- Cao, X., Chen, J., Imura, H., Higashi, O., 2009. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sens. Environ.*, 113(10), pp. 2205–2209.
- Chen, Z., Zhang, Y., Guindon, B., Esch, T., Roth, A., Shang, J., 2013. Urban land use mapping using high resolution SAR data based on density analysis and contextual information. *Can. J. Remote Sens.*, 38(6), pp. 738–749.
- Gong, P., Howarth, P.J., 1990. The Use of Structural Information for Improving Land-Cover Classification Accuracies at the Rural-Urban Fringe. *Photogramm. Eng. Remote Sensing*, 56(1), pp. 67–73.
- Hall-Beyer, M., 2017. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sens.*, 38(5), pp. 1312–1338.
- Hongmei Zhao, Xiaoling Chen, 2005. Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. In: *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. IEEE.* pp. 1666–1668.
- Karathanassi, V., Iossifidis, C., Rokos, D., 2000. A texture-based classification method for classifying built areas according to their density. *Int. J. Remote Sens.*, 21(9), pp. 1807–1823.
- Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I., 2016. Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 9(5), pp. 1830–1840.
- Li, N., Liu, F., Chen, Z., 2014. A texture measure defined over intuitionistic fuzzy set theory for the detection of built-up areas in high-resolution SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 7(10), pp. 4255–4265.

Pesaresi, M., Ehrlich, D., Caravaggi, I., Kauffmann, M., Louvrier, C., 2011. Toward global automatic built-up area recognition using optical VHR imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 4(4), pp. 923–934.

Pesaresi, M., Gerhardinger, A., Kayitakire, F., 2008. A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 1(3), pp. 180–192.

Sexton, J.O., Song, X.P., Huang, C., Channan, S., Baker, M.E., Townshend, J.R., 2013. Urban growth of the Washington, D.C.-Baltimore, MD metropolitan region from 1984 to 2010 by annual, Landsat-based estimates of impervious cover. *Remote Sens. Environ.*, 129, pp. 42–53.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.*, 117, pp. 162–176.

Xu, H., 2008. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.*, 29(14), pp. 4269–4276

Yuan, H., Van Der Wiele, C.F., Khorram, S., 2009. An automated artificial neural network system for land use/land cover classification from landsat TM imagery. *Remote Sens.*, 1(3), 243–265.

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.*, 24(3), pp. 583–594.

Zhang, J., Li, P., Xu, H., 2013. Urban built-up area extraction using combined spectral information and multivariate texture. *In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS. IEEE*, pp. 4249–4252.

Zhou, W., Troy, A., 2008. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *Int. J. Remote Sens.*, 29(11), pp. 3119–3135.