A novel band selection architecture to propose a built-up index for hyperspectral sensor PRISMA

Shishir Gaur

Indian Institute of Technology BHU

Nilendu Das

Indian Institute of Technology BHU

Rajarshi Bhattacharjee (✉ rajbhatt78645@gmail.com)

Indian Institute of Technology BHU

Debanirmalya Patra

Indian Institute of Technology BHU

Anurag Ohri

Indian Institute of Technology BHU

Abstract

Processing of hyperspectral remote sensing datasets poses challenges in terms of computational expense pertaining to data redundancy. As such, band selection becomes indispensable to address redundancy while preserving the optimal spectral information. This paper proposes a novel architecture using Genetic Algorithm (GA) optimizing technique with Random Forest (RF) classifier for efficient band selection with Hyperspectral Precursor of the Application Mission (PRISMA) dataset. The optimal bands are BLUE (λ=492.69nm), NIR (λ=959.52nm), and SWIR 1 (λ=1626.78nm). This paper also involves an application of the selected bands to accurate identify and quantify built-up pixels by means of a new spectral index named Hyperspectral Imagery-based Built-up Index (HIBI). The proposed index was used to map built-up pixels in six cities around the world namely Jaipur, Varanasi, Delhi, Tokyo, Moscow and Jakarta to establish its robustness. This analysis shows that the proposed index has an accuracy of 94.02%, higher than all the other indices considered for this study. Moreover, the spectral separability analysis also establishes the efficiency of the proposed index to differentiate built-up pixels from spectrally similar land use or land cover classes.

Introduction

Land covers in urban areas tend to change more drastically over a short period than elsewhere because of incessant urbanization. During the time period from 1960−2018, the share of the global urban population increased from 33.61−55.27% (Zha et al. 2003; Mukherjee et al. 2020). Even the urban population of India has also increased from 17.97% in 1961 to 31.16% in 2011 and is expected to reach 40% by 2030 (Kaur and Luthra, 2018). The studies related to urbanization of late have been analyzed with renewed enthusiasm by urban planners, economists, and researchers. Urban expansion has been quantified using economic, demographic, and geographical approaches. The quantification of the urban measurements with an economic and demographic perspective measures the change in the ratio of urban to the total population and the contribution of urban areas to gross domestic product (GDP). In comparison, the geographical aspect considers the spatial extent estimation of built-up infrastructure, including impervious surfaces such as concrete buildings and roads (Richards and Richards, 1999). Encroachment of natural lands will transpire due to the expansion of the urbanized areas. The transformation of these lands into the impermeable built-up area will devastate that area's hydrologic system, ecosystem, and biodiversity (Xu, 2007). It also changes the wind circulation, albedo effect, and surface temperature of the surrounding cities (Mallick et al. 2008). The measurement of built-up area has been significant because these land types can indicate environmental quality and urban development (As-syakur et al. 2012). The main problem encountered while measuring or mapping urban areas is the assessment of change in land usage from non-residential to residential (Xu, 2008). The estimation of the built-up extent has been conventionally performed by measuring the spatial spread of the built-up footprint from ground surveying information (Mukherjee et al. 2020). The satellite imagery datasets have also been used for calculating the built-up extent and these imageries are advantageous due to their historical availability and large-scale spatial coverage (Guindon et al. 2004). For the remote sensing-

based indices calculation, the scientific community has primarily used the LANDSAT series satellite imageries due to their easy availability. The LANDSAT program was activated in 1972, and since its inception, these satellites have acquired multiple multispectral images that calculate the Earth's reflected solar energy falling in the visible and non-visible ranges of the electromagnetic spectrum (Mukherjee et al. 2020).

The process of built-up spread calculation can be performed using several remotely sensed datasets or different satellite imageries and spectral values based on the category of land use (Xu, 2008). These calculations and estimations can be analyzed with the help of several classification algorithms (Poyil and Misra, 2015; Rawat and Kumar, 2015). Among the numerous classification techniques, the index-based thresholding method has been frequently applied by researchers because of its computational efficiency and ease of implementation (Zha et al. 2003; He et al. 2010). The urban land-use class of the Pearl River Delta of China has been classified by Chen et al. (2006) using multiple remote sensing-based indices with high accuracy. For mapping the bare land and built-up in urban areas, several indices have been used in various studies, such as Urban Index(UI)(Kawamura et al. 1996), Normalised Difference Built-up Index (NDBI) (Zha et al. 2003), Normalised Difference Bareness Index (NDBaI) (Zhao and Chen, 2005), Index-based Built-Up Index (IBI) (Xu, 2008), Enhanced Built-Up and Bareness Index (EBBI) (As-syakur et al. 2012), Dry built-up index (DBI) (Rasul et al. 2018), and powered B1 built-up index (PB1BI) (Mukherjee et al. 2020). Indices like UI, NDBI, and EBBI are based on the high-speed mapping of bare land or built-up areas. Nevertheless, these indices are incompetent enough to verify the proper distribution among the built-up and the bare land classes (He et al. 2010; Ukhnaa et al. 2019). Some researchers stated that this inability, because of the severe difficulty of the spectral response patterns to built-up areas, vegetation, and bare land, predominantly in terms of the pixel groupings in areas with heterogenic objects (He et al. 2010). All these built-up indices have been derived using LANDSAT imageries. The LANDSAT-series satellite(s) are multispectral sensors having only a few bands (Loveland and Irons, 2016). As compared to the multispectral data, the hyperspectral data provide more substantial information. The greater level of spectral details provides better prospects to analyze the Land use/ Land cover (LU/LC) pattern (Jarocińska et al. 2022). The band combination which has been best suited for the built-up delineation has been obtained using a genetic algorithm (GA) based optimization technique using random forest (RF) as a classifier (Nagasubramanian et al., 2018).

This study uses hyperspectral data to derive a new index for built-up area delineation. The PRISMA sensor dataset has been used in this analysis. This sensor was built by the Italian space agency and launched in 2019. This hyperspectral sensor consists of approximately 250 bands in a spectral range of 400−2500 nm. PRISMA images have a spatial resolution of 30 meters, similar to that of LANDSAT imageries (Vangi et al. 2021). The index has been named the 'Hyperspectral Imagery-based Built-up Index (HIBI),' which can properly delineate the built-up features and distinguish between built-up and non-built-up features. The novelty factor in this work is that for the first time the GA-based RF classifier has been applied to the PRISMA datasets. The optimal bands have also been figured out which can be used by research community for urban pattern indentification in future.

This index has been tested by creating classified maps of six cities in the world. The HIBI mapping outcomes have been compared to the results of several other existing built-up indices. The proposed index has also been compared with a machine learning (ML) and deep learning (DL) classifier.

Materials And Methods

Study area description

For analysing the performance classification of the given index HIBI, three cities have been chosen from India, and three other cities have been chosen from outside India. The study region selected from India is Delhi, Jaipur, and Varanasi. These cities have a high population density. Delhi is the national capital of India. It is the second most populous city in India. The city of Jaipur is the state capital of Rajasthan and the most populous city of the state. It is also the tenth most populous city in India (CensusInfo India, 2011). Varanasi is known as the spiritual capital of India and one of the most famous cities in the world (Garg et al. 2020). The cities that have been selected outside India are Tokyo, Moscow, and Jakarta. The city of Tokyo is known as the most populous city in the world. Moscow is the second most populous city in Europe. Jakarta is the national capital of Indonesia, and it also features in the top hundred most populous cities of the world (UN, DESA, PD 2014; Hui et al. 2017). The location map of the study site has been drawn in Fig. 1.

Datasets used

The remote sensing datasets comprise PRISMA satellite images. All the datasets used in this analysis have been Level-2D products. These datasets have been atmospherically corrected and geocoded (Vangi et al. 2021). Except for the proposed index, all the other indices have been estimated from LANDSAT-8 imageries. The PRISMA and LANDSAT-8 images are from April 2021.

Methodology adopted for the study

The built-up indices UI, NDBI, IBI, EBBI, DBI, and PB1BI, have been computed for the preprocessed LANDSAT-8 scene. The output images (represented in the form of maps) have been generated for each built-up index applying the optimum threshold. Then the proposed built-up index (HIBI) is estimated using the PRISMA dataset, and a comparison is drawn with all the other built-up indices. Then the comparison of the HIBI is made with a machine learning (ML) classifier named Support Vector Machine (SVM) and a deep learning (DL) classifier known as Convolutional Neural Network (CNN).

Development of built-up index HIBI and its comparison with ML and DL techniques

The spectral curve has been made for LU/LC features using the PRISMA dataset for the six selected cities. Seven PRISMA bands situated in the EM spectrum zone between 400-450nm have not been considered for this analysis. These bands basically lie in the aerosol domain. The median reflectance

value of each feature class has been shown in the spectral graph. The spectral pattern of PRISMA bands of the study regions is represented in Fig. 2. Another major challenge while working with the hyperspectral dataset has been the selection of the best waveband combination. In this scenario, the major task has been to select the best possible band combination for built-up area delineation. So to mitigate this challenge, a genetic algorithm (GA) based optimization technique using random forest (RF) as a classifier has been implemented. GA has been known as a population-dependent stochastic search optimization method influenced by natural genetics and natural selection principles. Wavebands have been represented by long string bits and are known as chromosomes. A score has been assigned to each of these chromosomes using a fitness function (Goldberg, 2001). In this case, the fitness function evaluates how well these chromosomes (combination of bands) perform to discriminate between built-up and non-built-up regions. These chromosomes have evolved in consecutive generations using mutation, selection, and crossover genetic operators to explore the solution space unless the best solution has been achieved or end criteria have been encountered. Chromosomes for reproduction can be selected in many ways. One of the ways is to pick chromosomes pair from the whole population, which gives better fitness scores to execute crossover. Genetic details of the two chromosomes can be randomly combined using the crossover operator. Some part of a chromosome gets modified by the mutation operator, and it averts GA from selecting local optimal solutions (Nagasubramanian et al. 2018). It has been significant to select an appropriate fitness function carefully. In this analysis, the F1 score of the built-up class has been chosen to assess the performance of the classifier. F1 has been defined as the harmonic mean of recall and precision values (Powers, 2020). A good F1 score is also indicative of good classification performance. F1 score ranges from 0 to 1 (Fourure et al. 2021). F1 has been mathematically calculated using these formulas (Nagasubramanian et al. 2018):

$$Precesion = \frac{TruePositive}{TruePositive + FalsePositive}$$

1

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

2

$$F1score = \frac{2 \times Precesion \times Recall}{Precesion + Recall}$$

3

10-fold cross-validation has been conducted to evaluate the fitness of the classifier. The GA gives the best possible band combination. In this case, it selected Band 13 (λ = 492.69nm) from the BLUE region, Band 22 (λ = 562.73nm) from the GREEN region, Band 34 (λ = 669.81nm) from the RED region, Band 66 (λ = 959.52nm) from the NIR region, Band 129 (λ = 1626.78nm) from the SWIR1 region, and Band 197 (λ

= 2229.75nm) from the SWIR2 region. The flowchart for the GA-RF architecture for the best waveband combination selection is shown in Fig. 3.

The GA architecture has given a combination of six bands from six different EM regions for built-up delineation. But still, all six bands can not be used for built-up estimation. The careful analysis of the LU/LC spectral curve shows that the BLUE region depicts a high reflectance value for the built-up class compared to other classes. On the basis of this observation, the BLUE band has been kept as the first term in both the numerator and denominator of the proposed index. The urban pixels have a low reflectance value as compared to the barren land in the NIR band. It may be noted that the healthy vegetation reaches its reflectance crest in the NIR region, which makes it an automatic inclusion in the index to differentiate built-up and vegetation. However, the spectral response of dry vegetation and bare land peaks up in the SWIR1 band leading to a similar spectral footprint as the construction materials. Thus, these three bands have been used in the present study to considerably improve the extraction of built-up pixels by increasing the contrast between bare lands and settlements. Even with the water bodies (inland as well as the sea) in these regions, the built-up class has shown the separation. In the GREEN, RED, and SWIR2 regions, the spectral curve of the built-up class gets overlapped at several places with some of the other classes. All the observations, as mentioned earlier, have been taken into account. A generic settlement delineation index has been proposed to recognize the built-up pixels on the basis of a suitable threshold. The formula of the built-up index consisting of BLUE, NIR, and SWIR1 band has been shown as:

$$HIBI = \frac{BLUE(\lambda = 492.69nm) - NIR(\lambda = 959.52nm) - SWIR1(\lambda = 1626.78nm)}{BLUE(\lambda = 492.69nm) + NIR(\lambda = 959.52nm) + SWIR1(\lambda = 1626.78nm)}$$

4

In this analysis, GA has given which specific bands in the given regions need to be used to get the best result. Since the PRISMA image is a hyperspectral dataset, there will be multiple bands in each region. For example, this dataset contains 10–12 bands in the BLUE region of the EM spectrum. So unless an optimization algorithm has not been implemented, it will be impossible to select the best band out of the given10-12 bands. A similar scenario will also happen in other regions of the EM spectrum with the PRISMA data. In the NIR region, this dataset contains around 60–65 bands. Even the SWIR 1 and SWIR 2 regions also include a large number of bands. The GA-RF architecture has been implemented by using Google colab software (a cloud-based free python platform).

The thresholding technique can be helpful for assigning an upper cutoff (U) and lower cutoff (L) for a single pixel (Mukherjee et al., 2020). The pixels inside the range [L, U] have been delineated as settlement pixels, and the other pixels in the image have been marked as non-built-up pixels. For accurate classification using any index, the proper U and L bounds estimation must be done using statistical techniques from the sample training set. This technique must be non-trivial as the built-up sample pixels

do not emulate any parametric variations like Gaussian distribution. The bootstrapping thresholding technique has been applied in this study to overcome the difficulty of the non-trivial method.

The classification performance of the proposed index has been compared with other supervised classification algorithms along with ML and DL classifiers.

Spectral separability measurement

One of the commonly used methods for spectral separability is known as Jeffries-Matusita(JM) distance, and in this study, the JM distance method has been used. This method has been very reliable for spectral separability because it behaves like the probability of correct classification (Padma and Sanjeevi, 2014). The probability density of the spectral vectors, S_1 and S_2 for the bands (l = 1, 2,... ., L) has been denoted as p_l and $q_{l,}$ and the JM distance has been calculated (Ghiyamat et al. 2013) as:

$$JMDistance(S_1, S_2) = \sqrt{\sum_{l=1}^{L}[\sqrt{p_l} - \sqrt{q_l}]^2}$$

3

The JM distance ranges from 0 to 2, where 2 indicates the maximum separability (Rao et al. 2014). The LANDSAT-8 image has been chosen as the base image, and on that image, the 50 pure pixels have been chosen for each of the built-up, cropland, vegetation, bare soil, sandbar, and waterbody class. Then, these pure pixels group has been placed on the corresponding classified image generated from each index. Since the geo-coded location of the points has been the same, the points lie at the same location for all the index-based classified images. The spectral distance between built-up and all the other classes has been calculated for all index-based classified images. This procedure has been carried out in ERDAS Imagine software. The entire process has been reproduced for the HIBI index by taking the PRISMA datasets.

Creation of training and testing data

The training datasets have been implemented to generate the threshold interval window of the settlement-delineation indices and for training the Support Vector Machine (SVM) and CNN models. The testing dataset gives a reference for comparison of the delineated settlement results generated from the indices as well as from SVM and CNN classifiers. The datasets for training and testing have been generated from the respective LU/LC classification maps of the given cities. The datasets contain samples from both the built-up and non-built-up pixels. The same set of training pixels has been incorporated for threshold interval window generation for all the indices of settlement delineation and for the SVM as well as the CNN classifier training. Similarly, the same set of pixels can be utilized to evaluate and compare the performance of the classification between several techniques. The confusion matrix has been created, and the skeleton of the matrix is depicted in Fig. 4. Several parameters have been chosen for the accuracy measurements like Sensitivity, Specificity, Positive Prediction Value (PPV), Negative Prediction Value (NPV), Total accuracy, and Cohen's Kappa (κ) (Mukherjee et al. 2020).

Results

Performance of the built-up index HIBI

This work has analysed the settlement and non-settlement regions mapping in the considered study regions using HIBI transformation. The multispectral bands of the LANDSAT-8 satellite image have been used for making all the other considered indices except HIBI. The HIBI index has been compared with the other indices (UI, NDBI, PB1BI, EBBI, DBI, and IBI). The HIBI result has been compared with RF and CNN classifiers to estimate the effectiveness and accuracy. Additionally, the settlement and non-settlement coverage area of the study region has been tabulated.

Result of the spectral separability

The spectra separation between settlement and other non-settlement classes has been tabulated in Table 1. The HIBI index has the best spectral separability values between built-up and all the other non-built-up classes.

Table 1
JM statistics measuring separability between Built-up and Non-Built-up pixels

Indices	JM distance for the separability					
	Cropland-Built-up	Bare soil-Built-up	Inland waterbody-Built-up	Sandbar-Built-up	Vegetation-Built-up	Sea water body-Built-up
NDBI	0.71	0.81	1.42	0.79	1.29	1.43
UI	0.68	0.81	1.47	0.61	1.15	1.47
PB1BI	1.68	1.31	1.65	1.09	1.53	1.61
EBBI	0.73	0.83	1.52	0.81	1.41	1.49
DBI	1.04	0.89	1.56	0.91	1.45	1.51
IBI	0.69	0.86	1.49	0.59	1.12	1.48
HIBI	**1.92**	**1.85**	**1.95**	**1.91**	**1.96**	**1.93**

Threshold window for built-up indices

The bootstrap thresholding has been applied, and the threshold window, along with the range diagrams, has been depicted in Fig. 5.

The HIBI index has the highest built-up range percentage among all the indices. This shows that the HIBI index has been the most robust and dynamic index of all. It can classify the settlement and non-settlement regions more effectively. The false positive (FB) value will be less for HIBI as compared to other considered built-up indices. The built-up threshold range percentage has been tabulated in Table 2. The percentage here has been considered as the average percentage by combining all the six considered cities.

Table 2
Built-up range
percentage out of
the total range
window for the
considered indices

Indices	Built-up Range (in %)
NDBI	7.24
UI	5.04
PB1BI	8.12
EBBI	7.34
DBI	6.08
IBI	6.92
HIBI	30.18

Quantitative accuracy assessment of the built-up indices

For each index, the accuracy estimation has been performed by considering the testing pixels as a reference for the six above-mentioned cities. In Table 3, the accuracy parameters have been computed for all the considered indices along with the accuracies of RF and CNN classifiers. For all the accuracy parameters, the average value has been computed because all these parameters for each index, along with ML and DL algorithms, have been applied to the six considered cities.

Table 3
Measurement of the accuracy parameters for the testing pixels

Indices	Sensitivity	Specificity	PPV	NPV	Accuracy(%)	(κ)
NDBI	0.862	0.768	0.787	0.887	75.33	0.717
UI	0.837	0.579	0.588	0.869	64.91	0.617
PB1BI	0.901	0.778	0.751	0.908	85.55	0.819
EBBI	0.801	0.582	0.646	0.889	73.29	0.692
DBI	0.485	0.576	0.571	0.502	46.83	0.438
IBI	0.771	0.628	0.651	0.798	66.94	0.638
HIBI	**0.922**	**0.937**	**0.914**	**0.952**	**94.02**	**0.919**
ML technique						
SVM	0.941	0.965	0.964	0.963	95.12	0.937
DL technique						
CNN	0.970	0.982	0.978	0.962	97.27	0.951

The built-up and non-built-up regions have been estimated by using HIBI for the six considered cities. The results have been tabled in Table 4. GA algorithm has been implemented in this analysis to decide which individual bands need to be selected from each of the respective EM regions. The F1 score value for this algorithm has been estimated as 0.92.

Table 4
Built-up and non-built up area calculation for each city
using HIBI

Cities and its surrounding region		
Study area	Built-up (Km^2)	Non-built up (Km^2)
Jakarta	377.05	661.16
Tokyo	1427.03	2197.31
Moscow	1105.06	2507.17
Delhi	920.7	1485.58
Jaipur	134.53	1571.36
Varanasi	92.93	749.88

Qualitative assessment of the built-up indices

The visual comparison among all the indices along with RF and CNN in the form of classified images for one of the cities (i.e., Jaipur) has been shown. Figure 6(A) depicts the classified maps. From the maps, it can be seen that all the indices have been overestimating the built-up regions except HIBI. From the spatial distribution pattern images, it can be reckoned that UI shows maximum overestimation, followed by PB1BI and EBBI. Figure 6 (B) represents the Standard false colour composite (SFCC) and true colour composite (TCC) image of the study stretch.

Discussion

The urban indices based on remote sensing technology have been generally used to differentiate between bare soil and built-up regions. These indices exhibit a low level of accuracy because these land-use categories possess a high degree of homogeneity. However, the application of HIBI has been found to be very effective in discriminating between bare soil and built-up areas, which has been a significant limitation of several pre-existing indices. The bootstrapping thresholding has been applied to determine the index range. The HIBI has been created using hyperspectral data, and other indices have been generated by using multispectral data. This analysis also shows that hyperspectral images are better for pixel-based classification in comparison to multispectral images. From the HIBI images of Jaipur city, it can be clearly seen that this index can properly delineate and discriminate between built-up and bare soil classes. This index can also successfully distinguish between built-up and sandbar. The sandbar gets mixed with urban pixels, and this can procure the wrong classification result. The sandbar formation has been a common phenomenon for a river like River Ganga (Jain and Singh, 2020). This river has been the lifeline for Varanasi city, and one extent of this city has been situated near the river bank. So for the proper delineation of the city stretch, the sandbar needs to be categorized into non-settlement classes, and earlier indices have not been very capable of doing this. The HIBI index can be beneficial for delineating those cities which lie near the sandbar intruded river bank. The coastal cities having shorelines can also be appropriately demarcated by this index. In cities where the tree canopy covers the building, at those places, this pixel spectra-based index can produce an error-prone result because of the heterogeneous landscape (As-syakur et al. 2012). If the spatial resolution of the satellite imagery gets enhanced, it can retrieve better information in heterogeneous urban regions by capturing small-scale objects (Tran et al. 2011). The accuracy of HIBI has been almost at par with the SVM classifier and slightly less than CNN classifier. The SVM and CNN algorithm execution has been very challenging, and it requires substantial knowledge and skillset in computer programming. The proper functioning of the HIBI has been dependent on the appropriate choice of pixels used for training the index. This feature of proper selection has been associated with all the supervised algorithms for classification.

Conclusion

A new pixel-spectra-based remote sensing index has been presented and analysed for the delineation of the non-built and built-up regions of several cities (three in India and three outside India). The index has

been calculated by using the PRISMA images. Three bands finally, namely, Blue (λ = 492.69nm), NIR (λ = 959.52nm), and SWIR1 (λ = 1626.78nm), have been used for estimating the HIBI. The analysis indicates that the proposed index can be a more accurate alternative to map built-up pixels in comparison to the existing indices considered in this study. These three bands are the optimal bands and can be used by future researchres for built-up classification. This index provides better separability between the sandbar and built-up class than the other indices considered. The other indices except HIBI overestimate the built-up area for his region. Like all supervised classification methods, the performance of HIBI is subjected to the selection of training data, i.e., the training set needs to be selected cautiously to ensure optimal performance.

Declarations

Declaration of Interest

The authors report there are no competing interests to declare.

Data Availability Statement

LANDSAT-8 data had been accessed through the GEE platform. https://earthengine.google.com

(accessed on 9-10 May 2021). The PRISMA data had been analysed in EVNI 10.5 software (accessed on 12 May 2021).

Authors' Contribution

Dr. Shishir Gaur, Rajarshi Bhattacharjee and Nilendu Das conceptualized the work. Data preparation and analysis were done by Rajarshi Bhattcharjee and Debanirmalya Patra. The first draft of the paper was prepared by Nilendu Das. Finally, Dr. Anurag Ohri and Dr. Shishir Gaur revised the manuscript critically and approved this current version. All authors reviewed the manuscript.

References

1. As-syakur A, Adnyana I, Arthana IW, Nuarsa IW (2012) Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area. Remote Sens 4(10):2957–2970

2. Chen XL, Zhao HM, Li PX, Yin ZY (2006) Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. Remote Sens Environ 104(2):133–146

3. Fourure D, Javaid MU, Posocco N, Tihon S (2021), September Anomaly detection: how to artificially increase your f1-score with a biased evaluation protocol. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 3–18). Springer, Cham

4. Garg V, Aggarwal SP, Chauhan P (2020) Changes in turbidity along Ganga River using Sentinel-2 satellite data during lockdown associated with COVID-19. Geomatics Nat Hazards Risk 11(1):1175–1195

5. Ghiyamat A, Shafri HZM, Mahdiraji GA, Shariff ARM, Mansor S (2013) Hyperspectral discrimination of tree species with different classifications using single-and multiple-endmember. Int J Appl Earth Obs Geoinf 23:177–191

6. Goldberg DE (2001) Genetic Algorithms in Search, Optimization, and Machine Learning, MA USA

7. Guindon B, Zhang Y, Dillabaugh C (2004) Landsat urban mapping based on a combined spectral–spatial methodology. Remote Sens Environ 92(2):218–232

8. He C, Shi P, Xie D, Zhao Y (2010) Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. Remote Sens Lett 1(4):213–221

9. Hui C, Richardson DM, Visser V (2017) Ranking of invasive spread through urban green areas in the world's 100 most populous cities. Biol Invasions 19(12):3527–3539

10. Jain CK, Singh S (2020) Impact of climate change on the hydrological dynamics of River Ganga, India. J Water Clim change 11(1):274–290

11. Jarocińska A, Kopeć D, Kycko M, Piórkowski H, Błońska A (2022) Hyperspectral vs. Multispectral data: Comparison of the spectral differentiation capabilities of Natura 2000 non-forest habitats. ISPRS J Photogrammetry Remote Sens 184:148–164

12. Kaur RR, Luthra A (2018) Population growth, urbanization and electricity-Challenges and initiatives in the state of Punjab, India. Energy strategy reviews 21:50–61

13. Kawamura M (1996) Relation between social and environmental conditions in Colombo Sri Lanka and the urban index estimated by satellite remote sensing data. In *Proc. 51st Annual Conference of the Japan Society of Civil Engineers* (pp. 190–191)

14. Loveland TR, Irons JR (2016) Landsat 8: The plans, the reality, and the legacy. Remote Sens Environ 185:1–6

15. Mallick J, Kant Y, Bharath BD (2008) Estimation of land surface temperature over Delhi using Landsat-7 ETM+. J Ind Geophys Union 12(3):131–140

16. Mukherjee A, Kumar AA, Ramachandran P (2020) Development of new index-based methodology for extraction of built-up area from landsat7 imagery: Comparison of performance with svm, ann, and existing indices. IEEE Trans Geosci Remote Sens 59(2):1592–1603

17. Nagasubramanian K, Jones S, Sarkar S, Singh AK, Singh A, Ganapathysubramanian B (2018) Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. Plant methods 14(1):1–13

18. Powers DM (2020) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*

19. Poyil RP, Misra AK (2015) Urban agglomeration impact analysis using remote sensing and GIS techniques in Malegaon city, India. Int J Sustainable Built Environ 4(1):136–144

20. Rao DS, Prasad AVV, Nair T (2014) Application of Texture Characteristics for Urban Feature Extraction from Optical Satellite Images. Int J Image Graphics Signal Process 7(1):16

21. Rasul A, Balzter H, Ibrahim GRF, Hameed HM, Wheeler J, Adamu B, Najmaddin PM (2018) Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates. Land 7(3):81

22. Rawat JS, Kumar M (2015) Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. Egypt J Remote Sens Space Sci 18(1):77–84

23. Richards JA, Richards J (1999) Remote sensing digital image analysis, fifth ed. Heidelberg, Berlin. https://doi.org/10.1007/978-3-642-30062-2

24. Tran TDB, Puissant A, Badariotti D, Weber C (2011) Optimizing spatial resolution of imagery for urban form detection—The cases of France and Vietnam. Remote Sens 3(10):2128–2147

25. Ukhnaa M, Huo X, Gaudel G (2019), February Modification of urban built-up area extraction method based on the thematic index-derived bands. In *IOP Conference Series: Earth and Environmental Science* (Vol. 227, No. 6, p. 062009). IOP Publishing

26. Un DESA, Population Division (2014) World urbanization prospects: the 2014 revision. *United Nations Department of Economic and Social Affairs (UN DESA) Population Division, New York*

27. Vangi E, D'Amico G, Francini S, Giannetti F, Lasserre B, Marchetti M, Chirici G (2021) The new hyperspectral satellite PRISMA: Imagery for forest types discrimination. Sensors 21(4):1182

28. Xu H (2007) Extraction of urban built-up land features from Landsat imagery using a thematic oriented index combination technique. Photogrammetric Eng Remote Sens 73(12):1381–1391

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

30. Xu H (2010) Analysis of impervious surface and its impact on urban heat environment using the normalized difference impervious surface index (NDISI). Photogrammetric Eng Remote Sens 76(5):557–565

31. 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):583–594

32. Zhao H, Chen X (2005), July Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. In *International geoscience and remote sensing symposium* (Vol. 3, p. 1666)

33. **WEB REFERENCE**

34. Census Info India (2011) Population report for India.
http://www.dataforall.org/dashboard/censusinfoindia_pca/ (accessed 14 June 2022)
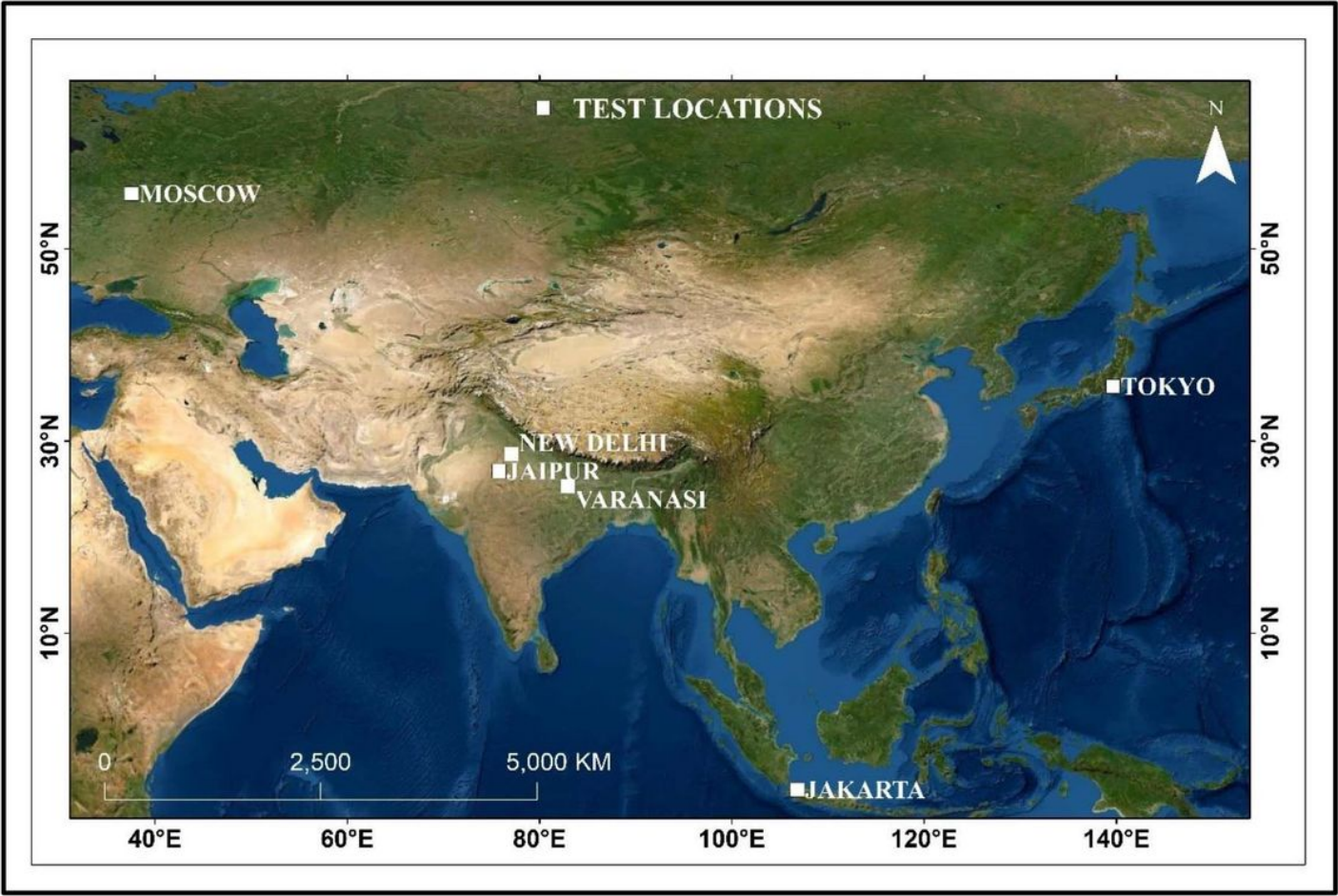
Figures

Figure 1

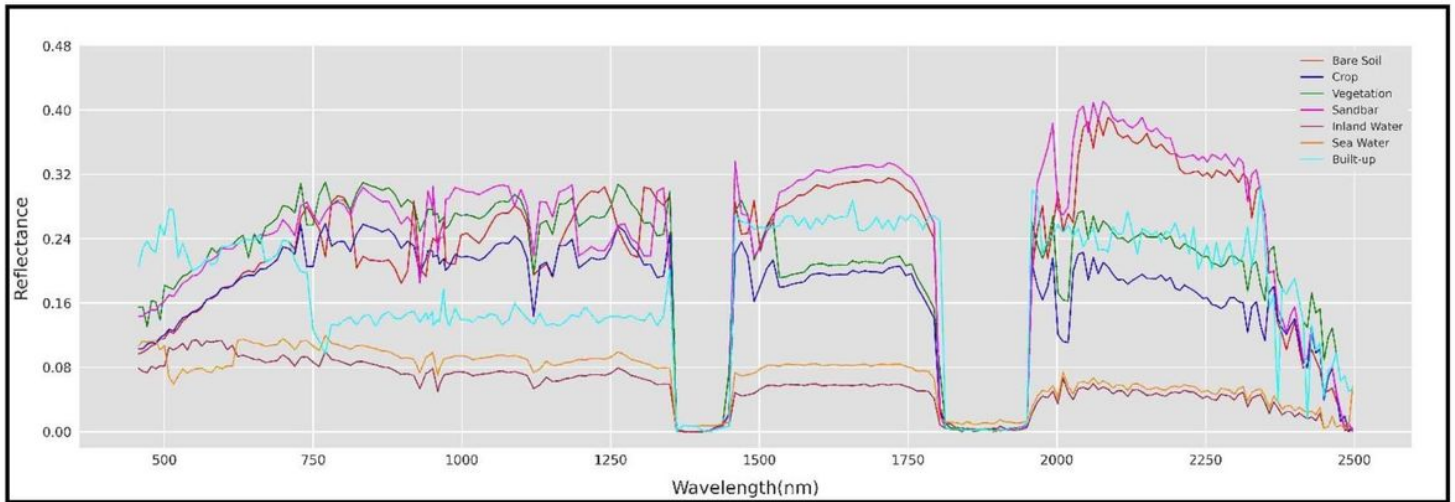**Figure 1**

Geolocation map of the study regions.

Figure 2

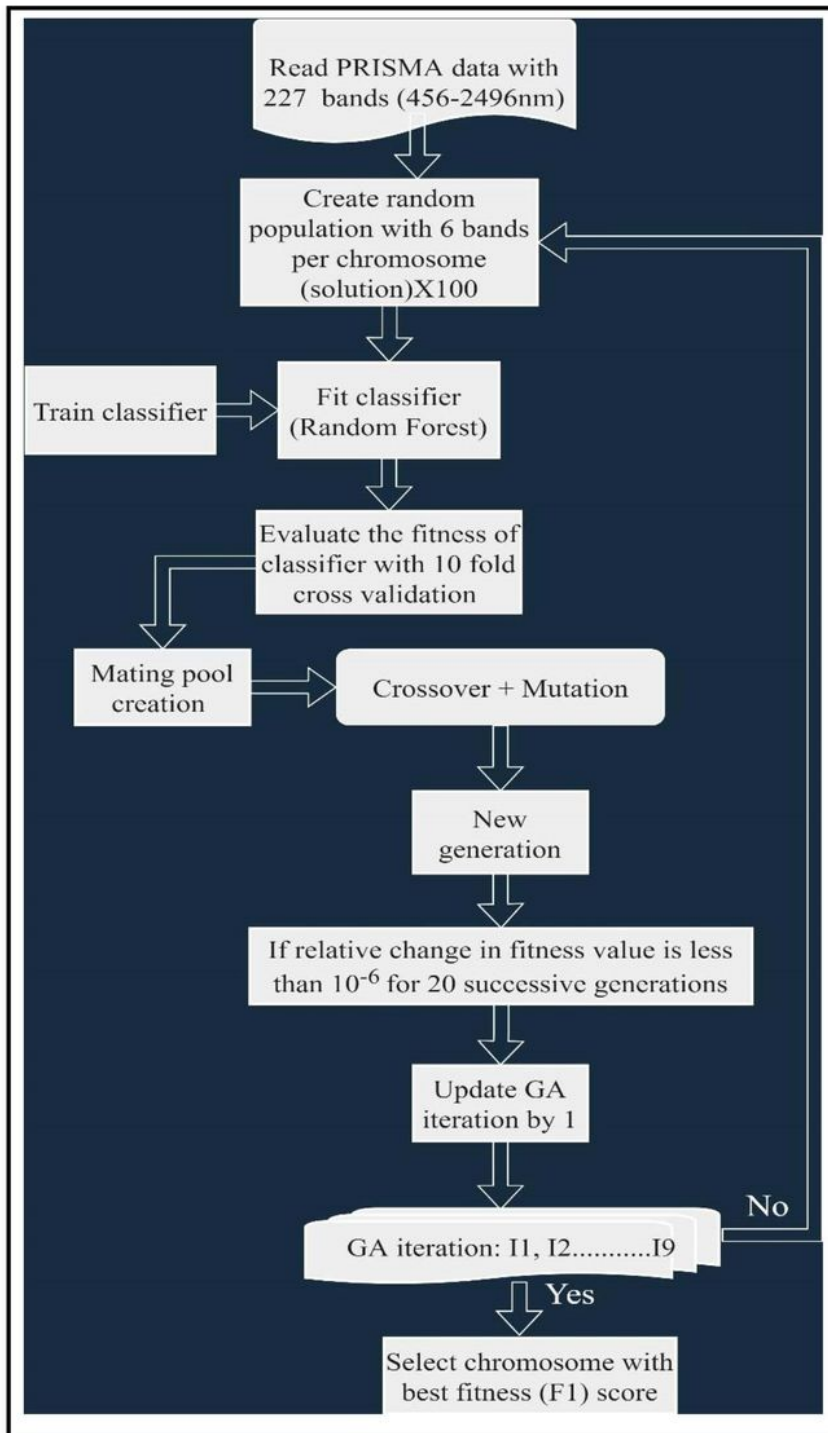**Figure 2**

Spectral profile for seven LU/LC classes

Figure 3

Figure 3

GA-RF architecture for optimal bands combination selection

Predicted pixels

	Built-Up	Non Built-Up	Total
Built-Up	True Positive (TP)	False Negative (FN)	P'
Non Built-Up	False Positive (FP)	True Negative (TN)	N'
Total	P	N	

Test pixels (vertical axis label)

Figure 4
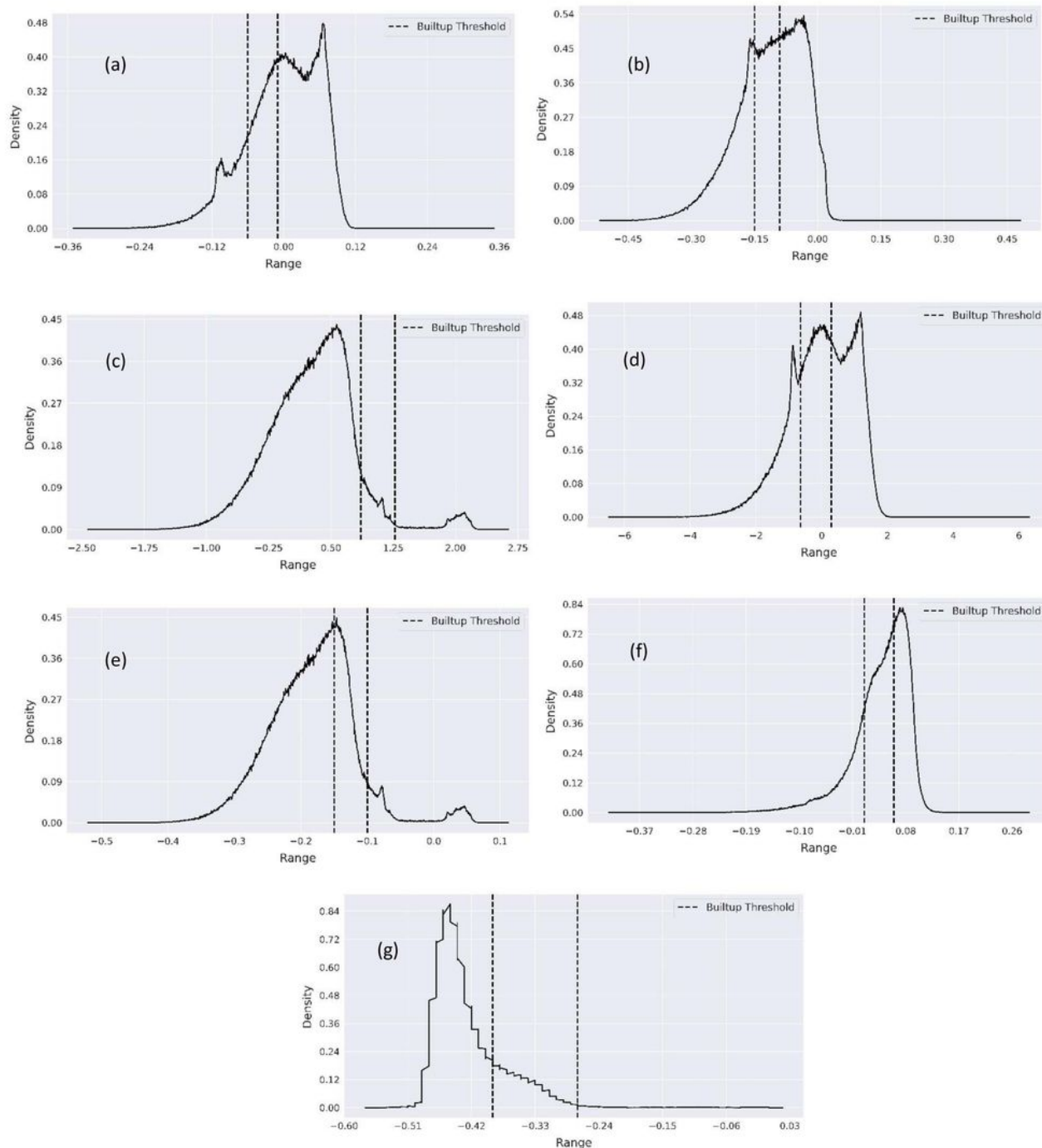
Figure 4

Skeleton of the confusion matrix

Figure 5

Figure 5

Threshold window of built-up indices using bootstrap threshold (a) NDBI (b) UI (c) PB1BI (d) EBBI (e) DBI (f) IBI (g) HIBI
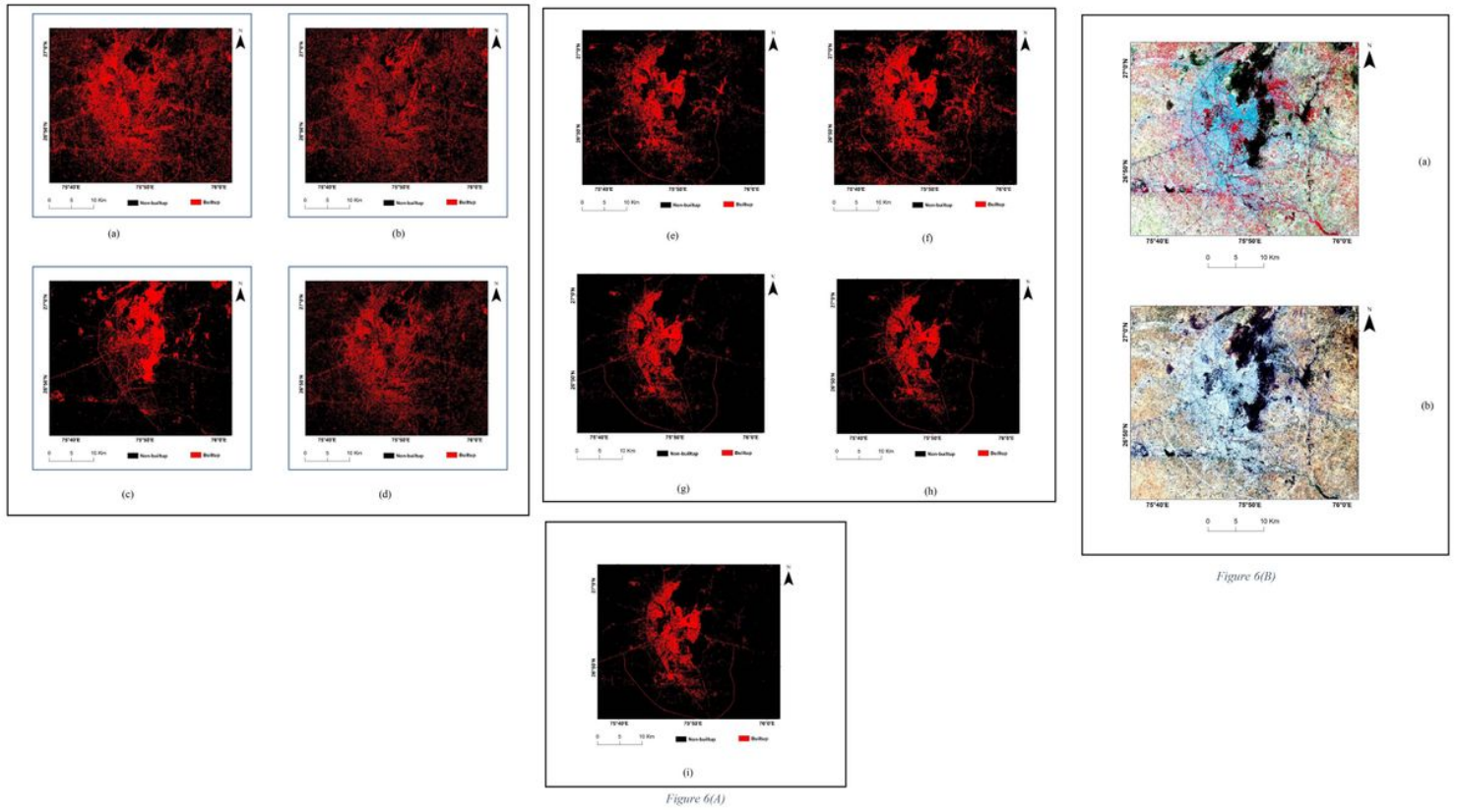
Figure 6(A)

Figure 6(B)

Figure 6

6 (A): Classified built-up and non-built-up maps of the study region using built-up indices and advanced classifiers (RF and CNN); (a) NDBI (b) UI (c) PB1BI (d) EBBI (e) DBI (f) IBI (g) HIBI (h) RF (i) CNN

6 (B): (a) SFCC and (b)TCC image