



Remote Sensing of Land

Homepage: www.gathacognition.com/journal/gcj1
<http://dx.doi.org/10.21523/gcj1>



Original Research Paper

Automated Building Extraction using High-Resolution Satellite Imagery through Ensemble Modelling and Machine Learning



Soumya K. Das¹, Prakash P. S.², Bharath H. Aithal^{2*}

1. Center for Land Resource Management, Natural Resource Management, Central University of Jharkhand, Ranchi-835 205, India.
2. RCG School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur, Kharagpur-721 302, India.

Abstract

Building extraction has been a challenging task due to complex structures and features of various land use with matching spectral and spatial attributes in a satellite data. We attempted to extract building as features using machine-learning algorithms such as Support Vector Machine (SVM), Random Forests (RF), Artificial Neural Network (ANN) and Improved Ensemble Technique as Gradient Boosting. The techniques used increases their classification accuracies using spectral properties as well as indices such as Normalized Difference Vegetation Index (NDVI) as attributes. Extracted results through various methods, performance of three different machine learning such as Ensemble method, RF and SVM are applied and results are analyzed for their behavior in different building distribution. Different algorithms showed variations in accuracies and performance in different built-up conditions. Ensemble algorithm performed very well in all conditions followed by RF and SVM performed better in coarse resolution, while ANN performed better in high resolution and overall accuracies of all algorithms increased with better spatial resolution. Ensemble algorithm showed relatively efficient performance in regions with extensive heterogeneous features. These analyses can help to provide quantitative data for various stocktaking analysis and city managers for better administration capabilities.

© 2018 GATHA COGNITION® All rights reserved.

Article history

Received: 08 October 2018;
Accepted: 13 November 2018

Keywords

Building extraction;
Ensemble modeling;
NDVI;
Model stacking.

Editor(s)

Suresh Kumar

1 INTRODUCTION

Building extraction from the satellite imagery has been a very challenging application in geographical urban data development. In the recent years, increased availability of quality (both spectral and spatial) remote sensing satellite imagery have filled up the void in technology and have given researchers the ability to develop algorithms for building extraction from the very High-resolution satellite imagery. Spatial resolution forms a very important part in the identification of buildings to perform building extraction and detection.

Building detection using the satellite imagery has significant applications in land use change mapping, urban planning, disaster risk assessment, urban space

use mapping. Researchers across the globe have developed many indices for various feature extraction including urban, land cover, vegetation, water bodies. Spectral-based indices (Xu, 2008) are the most commonly used for urban structure extraction (He et al., 2010). Examples include Normalized Difference Built-up Index (NDBI), Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI) and Modified Normalized Difference Vegetation Index (MNDVI), etc., (Ramachandra et al., 2012; Ramachandra et al., 2016; Bharath et al., 2018). But, spectral indices couldn't solely be used for extraction of buildings because of a significant amount of spectral mixing and heterogeneity of pixels of various land cover

* Author address for correspondence

RCG School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur, Kharagpur- 721302, West Bengal, India.
Tel.: +91 8861551265
Emails: dassoumyakanta96@gmail.com (S. Das); psprakashgi@gmail.com (P. Prakash); bhaithal@iitkgp.ac.in (B. Aithal -Corresponding author).

<http://dx.doi.org/10.21523/gcj1.18020103>

© 2018 GATHA COGNITION® All rights reserved.

types of urban scenario mainly, water bodies, built-up, barren land, as a result, reduces the accuracy of extraction (Sinha et al., 2016). To overcome these issues, two methods have been used: (i) classification involving various algorithms at pixels and object levels and (ii) enhanced spectral indices for buildings and then thresholding.

In machine learning, spectral indices can be used as attributes in identifying the pixels containing the building in complement with the spectral bands. Hence, the algorithms are trained with the values in various attributes and their labels from which it learns those small intricacies of the relations between attributes increase the classification accuracies. In this study, NDVI is used to complement Principal Component Analysis (PCA) and generated multispectral bands are used for the machine learning process.

1.1 Machine Learning and Applications in Building Extraction

Last decade machine learning has been widely used in remote sensing, that involved an enormous amount of data. Machine learning is based on the process of designing of algorithms that can learn from the data provided by the user for a particular phenomenon under analysis (Lary et al., 2016). One of the applied areas of application is the extraction of the building edges and heights. Extraction of the building is a complex procedure because of homogeneity of the spectral properties with other urban objects in the satellite imagery. Machine learning approach for building extraction from Very High Resolution (VHR) satellite image is quite helpful as it takes various predictors variables for a final prediction of presence or absence of a building. Using machine learning many studies have tried to develop automatic building detection algorithms (Dahiya et al., 2013). These algorithms considered building as objects and then segmentation is used to create the image object using the spectral properties, texture, or the shape of the image object. The possibility of incorrect extraction of features was observed in a complex urban scene with decrease in accuracy.

Studies have also used deep convolution neural networks and residual networks along with guided filters for building extraction and showed outstanding performance in building extraction which results in high computational demand (Xu et al., 2018). Similarly, in Yuan (2016) simple convolution neural network was used for building extraction which showed higher accuracies in the developed urban environment but accuracies degraded when applied to rural and suburban areas which showed that machine learning can play an important role in building extraction process giving significant accuracies. In Aytekin et al. (2012), genetic algorithm for building extraction was used for a multispectral image with pan-sharpening, showed promising results in building extraction with complete rooftop extraction of the buildings. Also extracted some complete barren areas as building's rooftops and

is observed that it gave higher accuracies when both spatial and spectral properties were considered.

There has been various studies such as Chen et al. (2018) that used machine learning methods for classification of buildings. Support Vector Machine (SVM), Random Forests (RF) and AdaBoost were methods used to extract building data and reported that SVM provided a good accuracy. It is observed that SVM proved to be very good in building extraction but intermixing of pixels with surrounding urban areas was observed. Further, Dornaika et al. (2016) highlighted building detection using image segmentation and descriptors on aerial orthophoto. In this study segmentation of image is performed, after which region descriptors are provided for segments than machine learning is applied which showed higher accuracies. Region descriptors played an important role in the machine learning process as many features are attached to a single label, complementary to one another that are helping in getting greater accuracies. This study drew an inference as increasing attributes for labels the accuracies could be increased. Guo et al. (2016) have used Google Earth imagery using Convolution Neural Network (CNN) and AdaBoost methods in extraction of building. Studies including Das and Ghosh (2017) and Martin and Vatsavai (2013) showed that ensemble of different models or classifiers generally corrects the errors committed by a single classifier thereby, increasing the accuracies. Ensemble learning was also applied on Hyperspectral imageries that increased their classification accuracies (Wu et al., 2012).

Each of the machine learning algorithms such as (a) SVM, (b) Ensemble and decision tree based as RF, (c) Artificial Neural Networks (ANN), etc., have been extensively used specifically in the extraction of buildings.

- (a) SVM classifiers work by creating a separating hyperplane between the support vectors that the training pixels and separates two classes in the feature space. The optimization is done by maximizing the margin between the hyperplane and closest support vector or training samples. This works well even with small training data sets when classifying high dimensional data because it only takes into considerations those training samples that are close to the margin, called support vectors and thus has seen few applications in an urban context (Waske et al., 2009).
- (b) Random forests algorithm is an ensemble-based machine-learning algorithm that uses multiple classifiers to generate multiple outputs to vote for better outcome. This is comparatively a very recent development and application are still very few (Breiman, 2001, Gislason et al., 2006). RF uses each individual classifier and then uses each of them to train on a subset of training data and the result is the mode of the classes of the individual decision trees with the goal of decreasing the variance. The training samples are

divided by using the bagging technique and the split in each node is based on a test rule.

- (c) Artificial neural networks are based on the bioinformatics that mimics human brain structure and features to gain a perception of a phenomenon under study. ANN's ability to provide better outputs even when relationships between various variables are not known is an added advantage and is capable of establishing connections between predictor variables efficiently. ANN has been adopted in various studies in Built-up extraction (Lari and Ebadi, 2007; Patel and Mukherjee, 2015; Sahar et al., 2006).
- (d) Finally, ensemble modeling runs two or more related but different predictive models then fuses the predictions for a single score (Gunes et al., 2017). Model stacking is used in ensemble learning in order to improve the accuracy of prediction using base learners. Stacking is primarily an ensemble based meta-learning technique that combines information or predictions from different predictive models to generate a new model. Generally, the stacked model performs better from the individual learners due to its nature of smoothing and the ability to predict when the base models perform well and discredit it when it performs poorly. Studies have showcased ensemble based learning for prediction of missing data in a time series data from multiple learning modules (Das and Ghosh, 2017). Further studies like Puttinaovarat and Horkaew (2017) have used a fusion of multiple outputs from different classifiers for building extraction which resulted in a higher overall accuracy of extraction.

In this study, we have used gradient boosting as a method of ensemble learning, as proved to provide better accuracy. Based on above discussion we propose an ensemble based extraction algorithm in which, various algorithms or base learners i.e. SVM, RF, ANN are combined for a final extraction result. This method has the capability of correctly classifying misclassified pixels increasing the extraction and detection accuracy.

Thus, considering all these aspects, the main objective of this study is to find out the potential of three prominent machine learning algorithms e.g. SVM, RF, ANN in automatic building extraction for the urban area from very high-resolution satellite datasets.

We have been evaluated the potential of an ensemble method using stochastic gradient boosting for building prediction in the homogeneous and heterogeneous distribution of building structures and at different spatial resolution.

2 DATA BASE

In this study, high resolution and very high-resolution satellite data were used for building extraction. Building extraction in images is a difficult task due to non-linearity of the building structure and lack of homogeneous arrangement of

buildings. Spatial resolution is an integral part of the extraction process and is considered as one of the comparative assessment indicators for successful extraction. Resourcesat-2 LISS-IV data at 5.8m spatial resolution (Figure 1) was used to study and evaluate the behavior of algorithm at a medium spatial resolution with non-linearity and heterogeneity of building structures. Data from Worldview-2 (Figure 2) satellite captured at very high-resolution was provided by DigitalGlobe with a spatial resolution of 0.5m panchromatic, 2.5m multispectral used to study and evaluate the behavior of the algorithm at a highly spatial resolution with high non-linearity of buildings. WorldView-3 (Figure 3) images at a high spatial resolution of 0.4m panchromatic, 1.25m multispectral provided by DigitalGlobe was processed for the analysis as the region has buildings that are linear and homogeneous as tabulated in Table 1.

3 METHOD

Methods used for building extraction using machine and ensemble learning are follows the process shown in Figure 4.

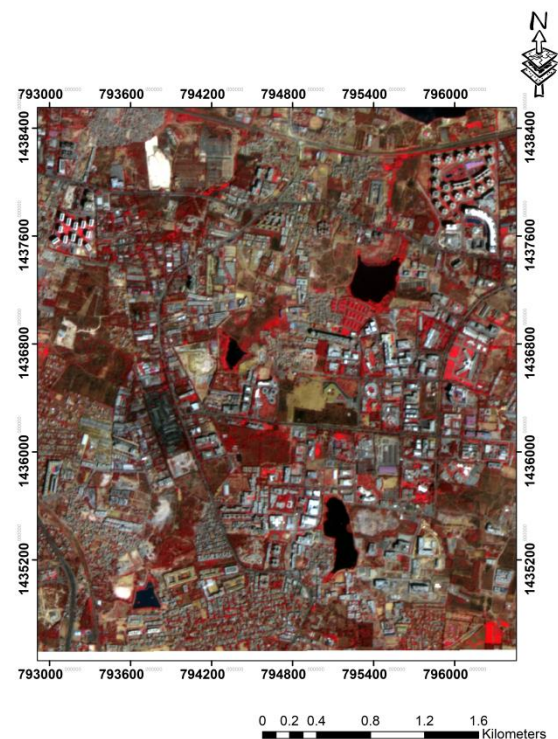


Figure 1: Resourcesat-2 LISS IV

Table 1. Data

Year	Satellite	Resolution
2014	Resourcesat 2 (LISS-IV)	5.8 m
2014	Worldview - 2	0.5 m PAN / 2.5 m MS
2014	Worldview - 3	0.4 m PAN / 1.25 m MS



Figure 2: Worldview-2

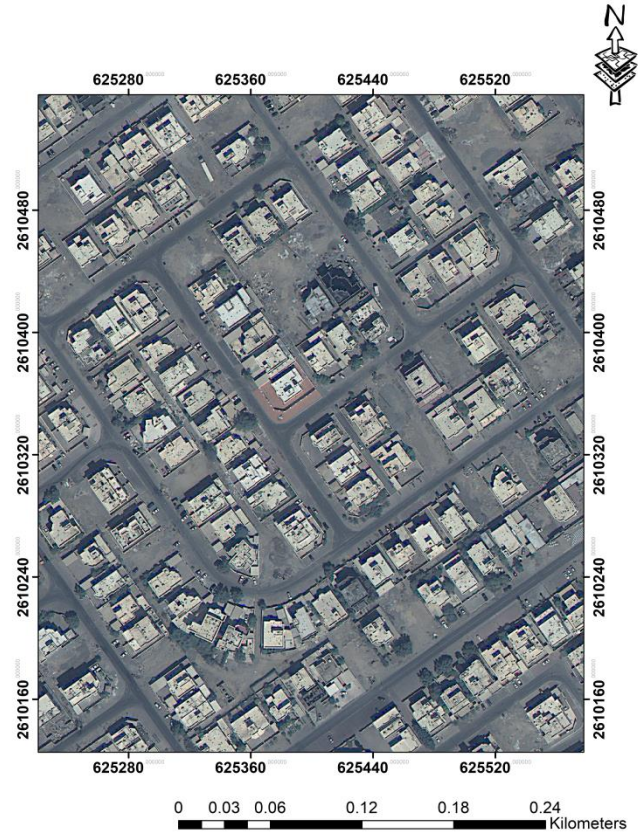


Figure 3: Worldview-3

3.1 Data Preprocessing and Preparation

Raw data sets are corrected for geometric and radiometric errors induced due to sensor distortions or due to the platform and atmospheric distortions. For radiometric corrections, histogram equalization was performed that maximizes the contrast of the data by applying a nonlinear contrast stretch that redistributes pixels so that there is equal number of pixels under each value within the range. Pan-sharpening was performed for the imagery to increase the spatial resolution for the extraction of buildings. Projective pan-sharpening was performed on the imagery as it gives much better

results as it preserves the color of the imagery as described in (Lindgren and Kilston, 1996).

Edge enhancement is performed on the image to enhance the edges in imagery with a sharp decrease and a sharp increase in intensity value of pixels to increase the probability of building extraction. This pan-sharpened and edge enhanced image are used to calculate the spectral indices and principal component. Only edge enhancement was performed on the LISS-IV imagery while both edge enhancement and pan-sharpening of multispectral bands of WorldView-2 and WorldView-3 were performed using their respective panchromatic bands of 0.5 m and 0.4 m resolutions.

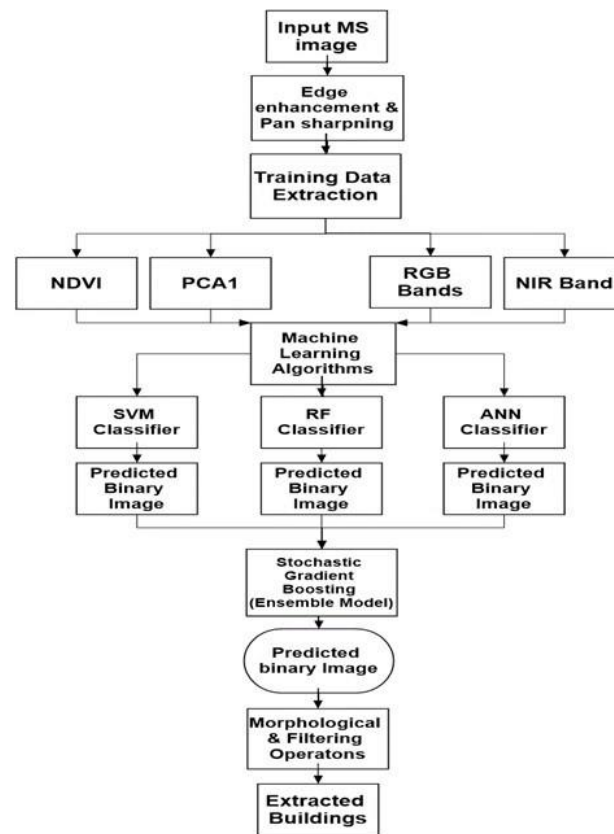


Figure 4: Process of building extraction

3.2 Spectral Indices and PCA of Satellite Imagery

NDVI (Bhandari *et al.*, 2012) is calculated using the pan-sharpened and edge enhanced imageries. NDVI has been used specifically for monitoring of vegetation and land cover studies. It basically measures the amount of chlorophyll present in the plants. The plant absorbs the infrared region of the spectrum than in the red region so the contrast between the two is used to build an index as described in equation (1). Different land cover has variations in NDVI values. Vegetation has very high value that decreasing with decreasing chlorophyll content. Buildings also have specific NDVI value based on the scene characterizes, for the study area it varies between -0.02 and 0.6. NDVI is calculated for the dataset which is used as an attribute for input data to machine learning algorithm for extraction of buildings.

$$NDVI = (NIR - RED) / (NIR + RED) \quad (1)$$

PCA (Estornell *et al.*, 2013) was performed based on the Infrared (IR), Red, Green and Blue (RGB) which gives various principal components as with decreasing variance and decreasing content of information. First principal component was selected as it has high variance.

3.3 Machine Learning Algorithms for Building Extraction from Very High-Resolution Satellite Imagery

After calculation of NDVI and PCA 1, each of the following algorithms was systematically used along with data input as building and non-building that are manually digitized. This manually digitized data is divided into a training dataset and test dataset (Verticale and Giacomazzi, 2008) in 7:3 ratio. Training data is used for training the algorithms and test dataset as validation data and is used for accuracy assessment of the algorithms.

3.3.1 Support Vector Machine Classification

SVM is a type of non-parametric supervised machine learning algorithms work by creating a separating hyperplane in a multi-dimensional feature space and that maximizes the orthogonal distance between hyperplane and the support vectors (Chapelle, 1998). Optimization is performed by maximizing the margin between the hyperplane and closest support vector or training samples. This works well even with small training data sets when classifying high dimensional data because it only takes into considerations those training samples which are close to the margin, which is, therefore, called support vectors. Training dataset consists of points $x_i \in R^N$, $i = \{1, \dots, l\}$. Where each data

point belongs to one of the two class given by class labels $y_i = +1$ or $y_i = -1$.

The algorithms work such that it draws decision boundaries or hyperplanes $w \cdot x + b$, where the decision rule function $w \cdot x_i + b \geq 0$ then, $y = +1$ and $w \cdot x_i + b \leq 0$ then, $y = -1$, where y is the label for the data point x_i . Optimization of equation (2) is used in generalizing the optimal hyperplane.

$$\min(\frac{1}{2}\|w^2\| + C \sum_{i=1}^l \xi_i) \quad (2)$$

Subject to $y_i(w^T \cdot x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

Here, ξ_i are positive slack variable and $C > 0$ is the regularization parameter or cost for any miscategorization errors.

If training datasets are linearly separable then the data points could be separated into two classes without error using a linear hyperplane. Generally, building structures are non-linearly separable cases with high non-linear relationships between the whole dataset and training datasets. Hence, as building distribution in imageries are more to near heterogeneous in nature, we use non-linear kernel functions. In which we exploit the SVM classifier radial basis function to map those data points in its feature space to higher dimensional feature space where the data points are separable. SVM linear classifiers find an optimal hyperplane using dot product between data point vectors (equation 3) and decision function shown in (equation 4).

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^T \vec{x}_j \quad (3)$$

$$f(\vec{x}) = \text{sign}(\sum_i \alpha_i y_i K(\vec{x}_i, \vec{x}) + b) \quad (4)$$

where, α_i is a Lagrange multiplier which is non-zero for support vector.

Kernel function as shown in equation (5) is then used to project every data point x_i to higher dimensional space via transformation is used.

$$K(\vec{x}_i^T, \vec{x}_j) = \phi(\vec{x}_i) \phi(\vec{x}_j) \quad (5)$$

where, $w^T \cdot \phi(x_i) + b$ is the separating hyperplane in higher dimensional space, it is transformed to a decision function equation (6), non-linear decision function, after optimal α_i is found, a sample x_i can be classified by finding out the side of hyperplane it lies in the new feature space (Waske et al., 2009).

$$f(\vec{x}) = \text{sign}(\sum_i \alpha_i y_i (\phi(\vec{x}_i) \cdot \phi(\vec{x})) + b) \quad (6)$$

As per the literature, there are mainly two groups of kernel function: Polynomial kernel and Radial basis kernel. In this study radial basis function (equation (6)) is used because it could classify points when the relationships between the labels of the class and attributes follow a non-linear pattern. It could be helpful in identification of buildings which generally follow a non-linear or heterogeneous arrangement. SVM and Radial basis function (RBF) (equation (7)):

$$K(x, \vec{x}_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2}) \quad (7)$$

Thus, the process of building extraction using SVM is performed using 5 major steps:

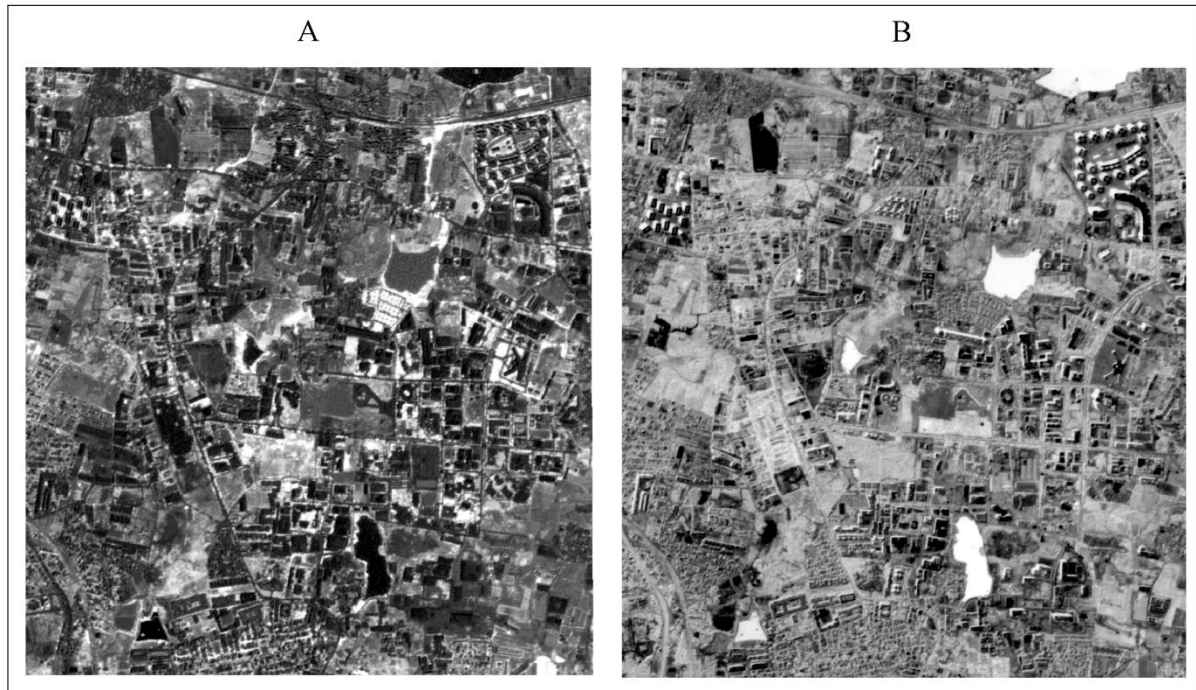


Figure 5. (A) NDVI and (B) PCA1

- The training data with values from various attributes are scaled and prepared for the input into the algorithm.
- To find the best-fit C and σ for the training dataset tuning was performed,
- RBF as kernel is used with the best C and σ on the training dataset to obtain a model.
- The model is used to predict the test dataset and accuracy is calculated as previously described
- After achieving a good accuracy, the model is used for prediction on the imagery. This also includes recoding the image as 0, 1, 0 for absence and 1 for presence.

3.3.2 Random Forests Classification

Random forests (Horning, 2010) machine learning algorithm is an ensemble-based supervised machine learning algorithm which uses multiple decision tree (DT) classifiers to create a forest and generate multiple outputs to vote for a final result using majority vote. RF works by first creating 'k' decision trees using the training dataset and the feature space by a technique called Boosting Aggregating or Bagging, which randomly selects and divides with replacement into bootstrap samples. Hence, there are samples those are not in the bootstrap samples are also called out-of-bag (OOB) samples, these samples are used for misclassification errors and variable importance called OOB error and then, it selects attribute features from given M attribute features for building decision trees. The best split at each node is chosen based on Gini impurity or information gain method, where, this is called Random subspace method. After the 'k' decision tree predicts their output, the prediction by the majority of the trees is chosen by RF as the final prediction.

If there are n samples and with feature vector x_i with labels y_i where, $\{i=1, \dots, n\}$, therefore, Data (D) = $\{(x_1, y_1), \dots, (x_n, y_n)\}$, Feature vector $(x_d) = (x_{d1}, \dots, x_{dn})$,

Many trees are built, in which each tree has a node with binary decision values, the division of tree continues until it has only one class and at each node feature (x_i) and threshold (a) , which are constant is chosen with minimum Gini criterion or impurity (Joelsson et al., 2005), if there is a subset of sample S , with two class C_1 and class C_2 then, variation $g(S_j)$, j is number of samples in S , as shown in equation (8),

$$g(S_j) = \sum_{i=1}^2 \hat{P}(C_i|S_j)(1 - \hat{P}(C_i|S_j)) \quad (8)$$

Where, $\hat{P}(S_j)$ proportion of S_j in S , $\hat{P}(C_i|S_j)$ is proportion of S_j in C_i . Variation $g(S_j)$ = Gini index = G under certain conditions. If each tree built is considered as weak classifier, then ensemble of weak classifiers is as complied by equation (9).

$$h = \{h_1(x), \dots, \dots, h_k(x)\} \quad (9)$$

$$\hat{h}(x) = \frac{1}{k} \sum_j \hat{h}_j(x) \quad (10)$$

Where, $\hat{h}_j$ (equation (10)) is a weak classifier or tree based on subset of S , for each tree we estimate the error on the unused data, averaging these errors give us the OOB (out-of-bag) error estimate.

3.3.3 Artificial Neural Networks Classification

ANN consists of a network of connections; in which the neurons work as computational units. In a feed-forward neural network, a bias node is added to each layer, a neuron that has a constant output and for each neuron from the input layer to the output layer, a weighted average of input is calculated then, activation function is applied which contains the neuron firing rules. There are various types of activation functions. They introduce non-linear properties into the network so, that they give it the ability to approximate more non-linear distribution (Lek et al., 1996) of data which is in case of building extraction. In this study in the feedforward network, we use a sigmoid activation function, with two input, a hidden layer and an output layer with two classes (building and non-building) and uses BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm (Ibrahim et al., 2014) for optimization of the network instead of back propagation.

Training dataset consists of input X_p , $p = \{1, \dots, i\}$. It calculates the weighted average (as shown in equation (11)) and sigmoid activation function (equation (12)) is used by the network and Y_j is the output.

$$Y_j = \sum_{i=1}^p (W_{ij}X_i + b) \quad (11)$$

$$Y_j = \frac{1}{1 + \exp(-V_j)} \quad (12)$$

The training datasets are prepared and scaled as the requirement of algorithm preferably its converted into the matrix, then a network is built, then the training data is feed into the network. Initially, sigmoid activation function is set to false, random weights are assigned as 0.1 with weight decay rate (decay) of $5e^{-4}$, and entropy is considered to be true for loss function. Best parameters are obtained using various combinations on the training data, hence builds a network model. Then it is used to predict the test dataset that is used for accuracy assessment. Once validated the model is used for predictions on the satellite imagery that extracts the building pixels to produce an image with building and non-building area.

3.4 Ensemble-Learning Based Model Stacking Approach for Building Extraction

Ensemble learning (EL) uses a group of weak learning algorithms to form a strong learner with greater accuracy than constituent base learners

(Choraś *et al.*, 2009). Ensemble modeling is running two or more different predictive models that predict different outputs, on which strong learner makes predictions for a single score. Two types of EL are generally used; One is Bootstrap Aggregating (Bagging) in which it trains each model with a random subset of training data with equal weight to each models output after that final prediction is the majority vote of all the predictions by all the models and another is Boosting (Gunes *et al.*, 2017), that builds model on the training data, then creates second model that tries to correct the error committed by first model, models are added until the training set is predicted correctly. We use Model Stacking in this study in order to improve the accuracy of prediction using base learners which are SVM, RF, ANN and to make predictions on the outputs of the base learners, we use stochastic gradient boosting an ensemble learner.

3.4.1 Advantages of Model Stacking Over Individual Classifiers and Stochastic Gradient Boosting Approach

Ensemble modeling approach first uses individual classifiers to make predictions on the training data, and then it uses the predictions from the base learners as input. The output of individual classifiers is then used to train the ensemble learner as illustrated in Figure 6. Generally, the stacked model performs better than the individual learners due to its nature of smoothing and the ability to predict when the base models perform well and discredit it when it performs poorly.

Further, after prediction by different machine learning algorithms, outputs are used as input for the second layer of the learning algorithm (Figure 6). Algorithm efficiently combines the predictions of the base learners to form a new model to produce higher accuracy. In the second layer of the learning algorithm generally models

that boost the accuracy of predictions for the complex relationship between input variables are normally used. Stochastic gradient method is one of such non-linear approximating algorithms that help in establishing a complex relationship between variables.

As shown in Figure 6, initial training dataset (X) has m rows or observation and n attributes, for instance, if we consider a number of models used in level 1 is M, M number of predictions and model's types are trained on the dataset. Then each model gives an output, y_i , Where $i = \{1, 2, \dots, M\}$.

This forms the training data of the second level consisting of m observations and M attributes. Finally, the second level model is trained on this data ($m \times M$) using gradient boosting, that gives the final predictions.

The aim of the Ensemble Learning Algorithms (Gradient boosting), in this case, is to find and minimize a loss function. We can describe loss as in equation (13) where $y_i = i^{\text{th}}$ target value, $y_i^p = i^{\text{th}}$ prediction, $L(y_i, y_i^p) = \text{loss function}$.

$$\text{loss} = \text{MSE} = \sum (y_i - y_i^p)^2 \quad (13)$$

The predictions should have minimum loss functions or error. It keeps on updating the predictions using gradient descent which moves forward by a learning rate until it finds predictions (equation (14)) with less MSE [Mean Squared Error].

$$y_i^p = y_i^p + \alpha * \delta \sum (y_i - y_i^p)^2 / \delta y_i^p \quad (14)$$

where, α is learning rate and $\sum (y_i - y_i^p)$ is sum of the residuals,

This type of boosting reduces both bias and variance and prevents over fitting of the model

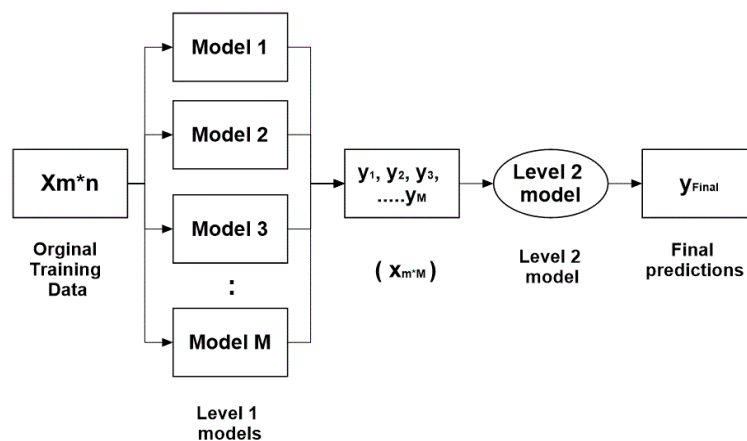


Figure 6. Ensemble modeling

on the data. Stochastic gradient boosting is a slightly changed version of the gradient boosting in which it uses bagging with boosting which induces randomness in function estimation. Random subsample of training data (without replacement) is used at each iteration for the full training data. This subsample is used to fit the base learner and then updates the prediction of the model for iteration. The parameters are tuned and set for optimal performance are loss function, number of iterations, depth of the trees, learning rate and sampling rate or bagging rate. As we now have the output of the three classifiers, the algorithm uses these outputs to build a model with optimal parameters, which is used for final building extraction by the ensemble.

3.5 Morphological and Cleaning Operations on Classified Imagery

Morphological and cleaning operations are performed on output image for removal of salt and pepper noise in the image (Kupidura and Jakubiak, 2009). Morphological operators are based on set theory that takes a binary image and a kernel as input and combines using set operators and gives a modified binary image. Open, dilate morphological operations are performed on the image. Clump operator was used to detecting patches of connected pixels. Finally, area filter is applied to the image to remove groups with less than five pixels to get a final extraction of buildings from the imagery. Morphological operations are performed on WorldView-2 and WorldView-3 imageries, while LISS-IV imagery is excluded due to loss of information.

4 ACCURACY ASSESSMENT AND MODEL PERFORMANCE

Accuracy assessment is performed to evaluate the correctly represented ground reality in the image (Foody, 2010). Confusion matrix is one of the most used methods of assessment of accuracy from which various types of matrices could be extracted for calculating accuracy and errors. This will also help us to access and analyze the performance of the model in different conditions using test data. The predictions of the model on test dataset will be used for generating confusion matrix, accuracy, kappa, sensitivity, specificity along with two metrics Branch Factor (BF) and Detection Percent (DP) (Lin and Nevatia, 1998).

DP is how many buildings get detected in the image by the algorithms and BF is the how many buildings are found improperly. Generating these metrics requires manual digitization and comparison as explained in equations (15) and (16).

Building detection (DP) rate:

$$DP = 100 * \left(\frac{TP}{TP+TN} \right) \quad (15)$$

$$\text{Branch Factor (BF)} = 100 \left(\frac{FP}{TP+FP} \right) \quad (16)$$

where, TP = building detected both manually and by the algorithm (sensitivity), TN = Building detected manually, but, not by an algorithm (specificity), FP= Building detected by the algorithm but, not manually. BF is how many buildings are detected erroneously.

4.1 Buildings Structure and Distribution effect on model performance.

Model performance is analyzed from simple to complex building structure and distribution: First, they are tested in two different images (LISS-IV and WorldView-2) of housing with a non-linear and heterogeneous pattern of distribution with two different resolutions. Secondly, these are tested in the WorldView-3 dataset with linear and homogeneous building distribution pattern where each building can be easily separated and medium mixing of spectral signatures from various urban structures.

5 RESULTS AND DISCUSSIONS

The targets of this study were: (a) to find the potential of the machine learning algorithms i.e. SVM, RF, ANN in building extraction from high-resolution satellite imagery and (b) to test and evaluate the ensemble learning technique in enhancing the performance of extraction of buildings. The complete statistical measures are shown in tables 2, 3 and 4 for various algorithms. Ensemble showed greater accuracy in three imageries. For the WorldView-2 imagery, SVM showed higher accuracy followed by ANN and RF. In WorldView-3, highest accuracy was showed by ANN followed by RF and SVM. This study showed that ANN performs significantly in uniform and linear building distribution on very high-resolution imagery. while RF showed good accuracy in the non-linear and heterogeneous distribution of urban buildings.

SVM worked very well in nonlinear building distributions as in WorldView-2 and optimal performance in other building distributions. It is observed that highest sensitivity is observed for RF for LISS-IV, while, lowest was shown by SVM. An equal degree of sensitivity is observed for both SVM and RF in WordView-2 imagery. The equal degree of sensitivity is observed for all three algorithms in case of WordView-3 imagery.

Gradient boosting algorithm showed improvement in accuracy of classification over single classifiers, it showed 1.58%, 0.2% higher accuracy and sensitivity, respectively than best performing algorithm in the LISS-IV imagery, 0.25% higher accuracy and 0.1% lower sensitivity than the best performing algorithm in WordView-2 imagery and 0.73%, 0.3% higher accuracy and

sensitivity than best-performing algorithm in WordView-3 imagery.

After applying morphological operators and cleaning filters on LISS-IV imagery it was observed that it could not retain building occupying less than 10 pixels or area of less than 336.4 square meters due to a coarser resolution. While, for others morphological operators clean the salt and pepper noise, refining the footprints of the buildings. In very high resolution nearly every building was correctly classified, but some parts of the roads were misclassified as buildings pixels.

It showed that algorithms mathematical and statistical way of working has an impact on their classification pattern for a building distribution. The variation of accuracies of algorithms is due to variable building distributions and spatial resolutions of the imagery. As expected higher accuracy in all conditions is observed for the ensemble algorithm as it maps complex relationship from the outputs of the base algorithms, thereby corrects the misclassified pixels by the algorithms. Building detection rate by ensemble method for high-resolution imagery observed was 91.30% and 95.05%, while a building getting detected erroneously (Branching factor) is 10.01% and 4.01%. The decrease is seen due to the linearity of house construction with less spectral mixing. While, SVM showed nominal behavior but, works very well in some complex parts of the imagery where other algorithms show discrepancy, because even if it can't find a linear decision boundary it can projects data points in multidimensional space to find a decision surface for this Radial basis function is used, thereby,

separating the points. However, there are some disadvantages of support vector machine, those are large number of calculations due to large number of support vectors and during testing, its speed of classifying points becomes very slow and many critical parameters are to be set. SVM method of classification produced salt and pepper noise, which could be removed by morphological operators and cleaning filters. Use of ANN showed no improvement in accuracy if the number of hidden nodes is increased after a certain number, using fewer features or attributes as input produced low extraction results, and learning rate is a significant effect on neural networks, finding an optimal learning rate is important for accuracy. It showed higher overall accuracy for high-resolution imageries. Finding optimal parameters for neural networks requires an extensive combination of parameters. While RF can actually distinguish confusing urban objects increasing their probability of detection with stable classification accuracy without loss of information hence, works well both in low and high-resolution imagery.

Building extraction results of different algorithms (figure 7- A, B, C, D) overlaid over imagery and final building extraction results for a subset of LISS-IV imagery (Figure 7). Final building extraction results for a subset of WorldView-2 imagery is as shown in figure 8.

Evaluation of detection rate in both images from Worldview 2 and Worldview 3 is shown in table 5. Comparable performance is tabulated graphically in figure 1.

Table 2. Statistical quantification using various algorithms for LISS-IV imagery

Machine Learning Models	Accuracy (%)	Kappa coefficient	Sensitivity	Specificity
SVM	87.13	0.73	0.84	0.91
RF	88.33	0.74	0.89	0.86
ANN	86.91	0.72	0.85	0.89
GDB	89.67	0.76	0.91	0.85

Table 3. Statistical quantification using various algorithms for WorldView-2 imagery

Machine Learning Models	Accuracy (%)	Kappa coefficient	Sensitivity	Specificity
SVM	93.09	0.84	0.94	0.92
RF	91.43	0.81	0.94	0.92
ANN	92.34	0.84	0.93	0.93
GDB	93.34	0.84	0.93	0.93

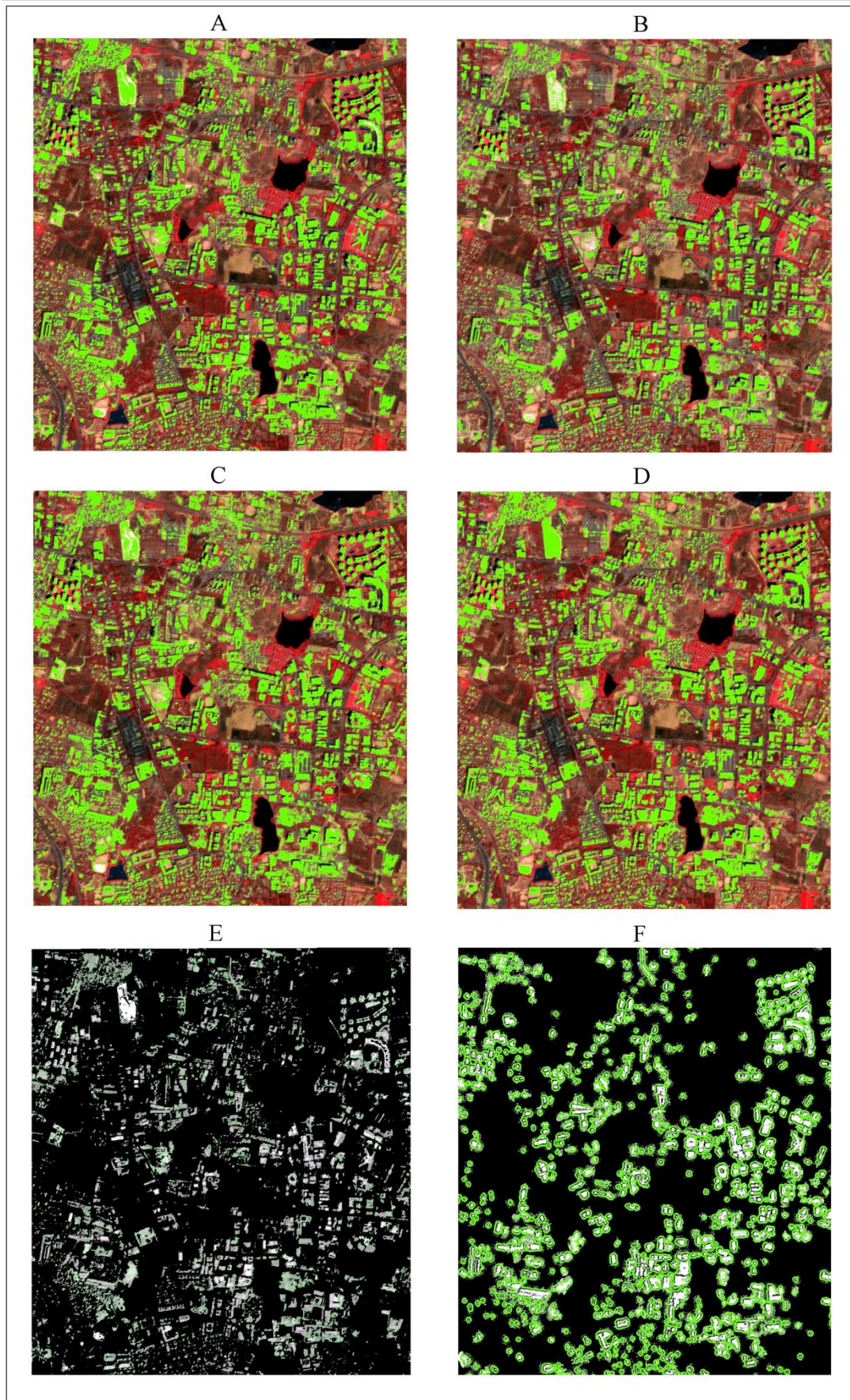


Figure 7. Building extraction for a subset of LISS-IV imagery: (A) SVM, (B) RF (C), ANN, (D) GB, (E) building extraction before and (F) after applying morphology operators

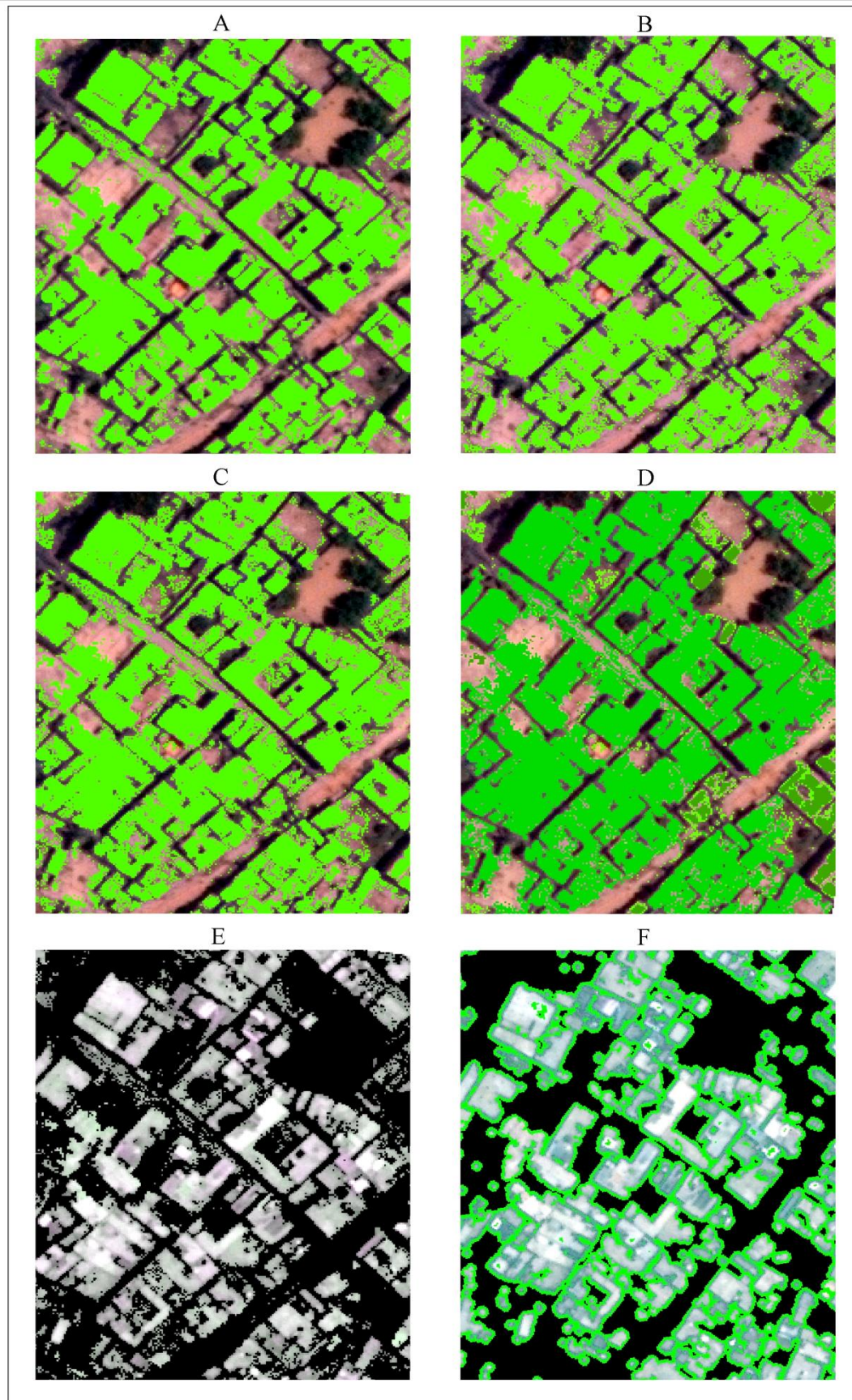


Figure 8. Building extraction for a subset of WorldView-2 imagery (A) SVM, (B) RF, (C) ANN, (D) GB, (E) Building Extraction before and (F) after applying morphology operators

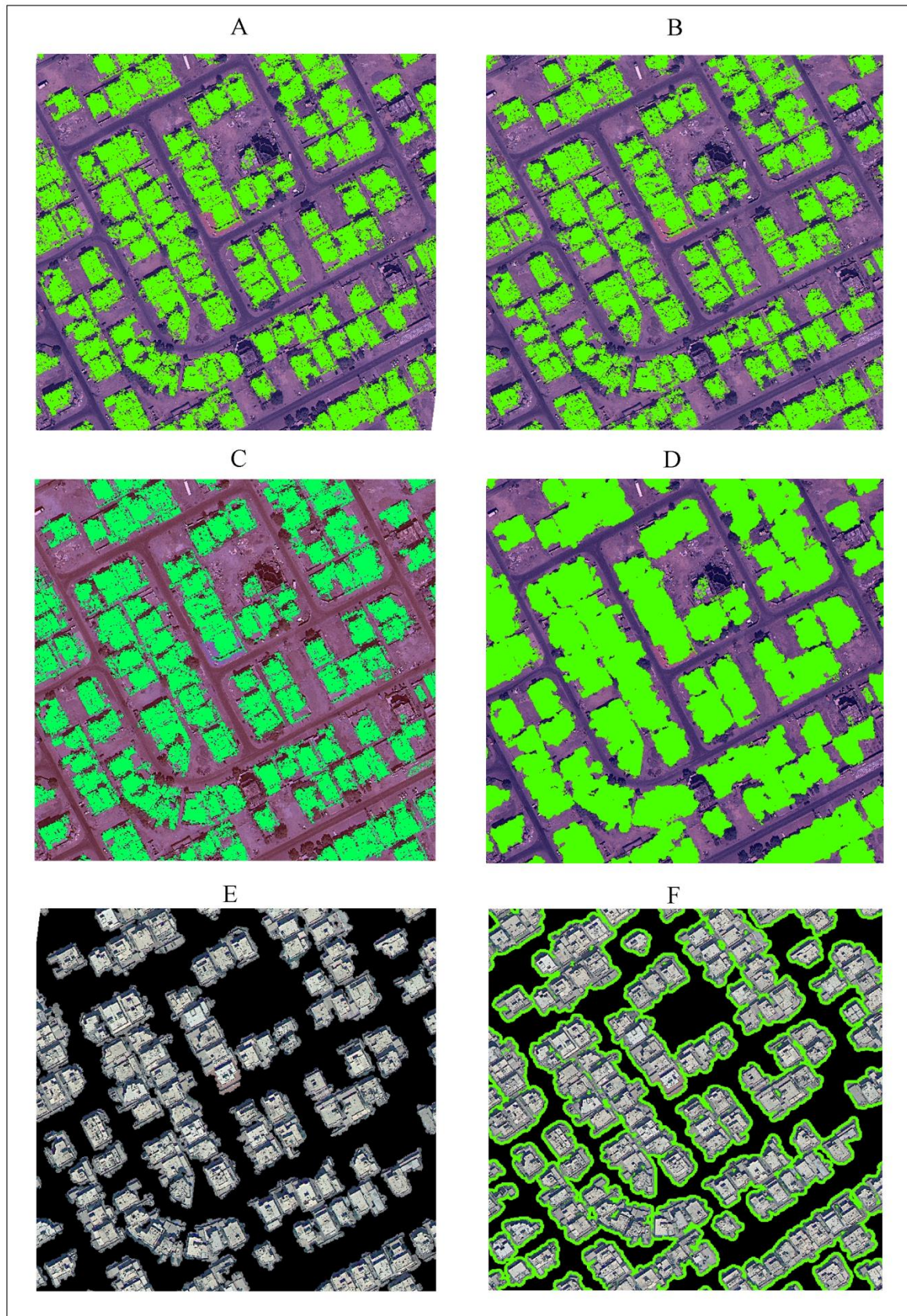


Figure 9. Building extraction for a subset of WorldView-3 imagery (A) SVM, (B) RF, (C) ANN, (D) GB, (E) building extraction before and (F) after applying morphology operators

Table 4. Statistical quantification using various algorithms for WorldView-3 imagery

Machine Learning Models	Accuracy (%)	Kappa coefficient	Sensitivity	Specificity
SVM	93.98	0.86	0.95	0.93
RF	94.83	0.89	0.95	0.95
ANN	95.05	0.90	0.95	0.94
GDB	95.78	0.87	0.98	0.94

Table 5. Evaluation of building extraction algorithms using two matrices detection rate and branching factor for high-resolution imagery

Image	TP	FP	TN	DP (%)	BF (%)
Figure 8(f).	63	7	6	91.30	10.01
Figure 9(f).	96	4	5	95.05	4.04

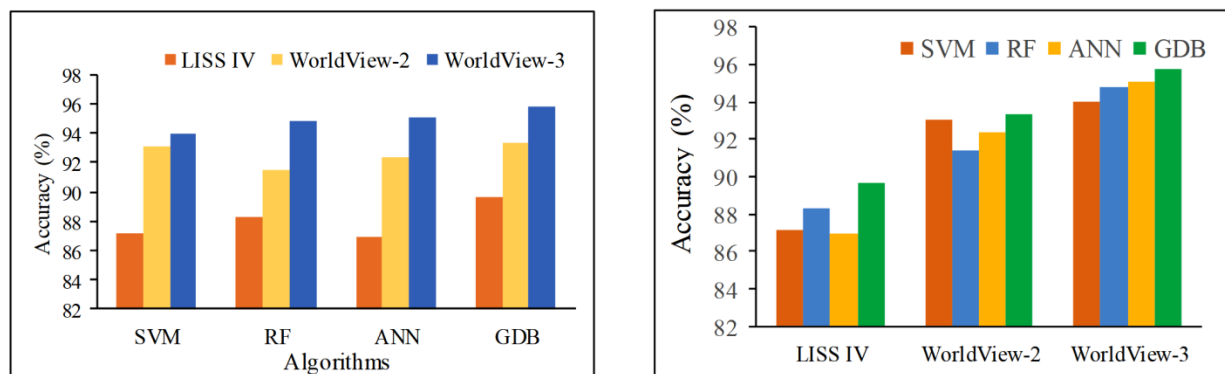


Figure 10. Comparable performance of the algorithms on different satellite imagery with different building distribution patterns

6 CONCLUSION

In this study, various types of machine learning algorithms and an ensemble learning technique were used for building extraction. These machine learning algorithms were successfully implemented for extraction of buildings with algorithms showing various amount of accuracy in different building distributions. Ensemble learning algorithm, which is the gradient boosting used for enhancing the classification accuracy, was successfully implemented for efficient building extraction from the high-resolution imagery. This is machine learning algorithm exploits various vegetation indices, PCA, spectral properties of the image which were fused into a single vector and where used for urban building extraction. It is observed that the overall performance and accuracy of the machine learning algorithms increased when the spatial resolution was increased with decreased misclassification rate. It is also observed that coarser resolution imagery such as in

LISS-IV imagery building occupying less than 10 pixels or area less than 336.4 square meters were not captured with higher misclassification rate and also few barren lands where misclassified as buildings. However, most of the buildings were captured properly with edges being detected and shape being maintained. While in very high-resolution imagery algorithms worked very well with high building detection rate and higher accuracy. Based on this study, RF showed good performance in coarse resolution imagery, while Support Vector Machine and Artificial Neural Network showed optimal performance in all conditions and good in high-resolution imagery while the ensemble algorithm was considered best for efficient extraction of buildings from imagery. Specifically, the individual algorithms are repeatedly misclassifies the particular part. Hence, it's observed that individual algorithms perform well in building extraction, while, ensemble learning could be used for increasing their accuracy of extraction.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to SERB, India, the Ministry of Science and Technology, Government of India, Ranbir and Chitra Gupta School of Infrastructure Design and Management and Sponsored Research in Consultancy Cell, Indian Institute of Technology Kharagpur for the financial and infrastructure support. We also thank Department of higher education, West Bengal and DST (W. B.) for financial support. We thank National Remote Sensing Centre (NRSC Hyderabad) for providing remote sensing data and other sources of satellite data.

ABBREVIATIONS

ANN: Artificial Neural Network; **BF**: Branch Factor; **DP**: Detection Percent; **MNDVI**: Modified Normalized Difference Vegetation Index; **NDBI**: Normalized Difference Built-up Index; **NDVI**: Normalized Difference Vegetation Index; **PCA**: Principal component analysis; **RF**: Random Forests; **SAVI**: Soil Adjusted Vegetation Index; **SVM**: Support Vector Machine; **SVM**: Support Vector Machine.

REFERENCES

- Aytekin, Ö., Erener, A., Ulusoy, İ. and Düzgün, Ş., 2012. Unsupervised building detection in complex urban environments from multispectral satellite imagery. *Int. J. Remote Sens.* 33, 2152-2177. DOI: <https://doi.org/10.1080/01431161.2011.606852>
- Bhandari, A. K., Kumar, A. and Singh, G. K., 2012. Feature extraction using normalized difference vegetation index (NDVI): A case study of Jabalpur City. *Procedia Technol.* 6, 612-621. DOI: <https://doi.org/10.1016/j.protcy.2012.10.074>
- Bharath, H. A., Chandan, M. C., Vinay, S., Gouri, H. A. and Ramachandra, T. V., 2018. Green to gray: Silicon Valley of India. *Journal of Environmental Management*, 206, 1286-1295. DOI: <http://dx.doi.org/10.1016/j.jenvman.2017.06.072>
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5-32. DOI: <https://doi.org/10.1023/A:1010933404324>
- Chapelle, O., 1998. Support vector machines for image classification. *Stage deuxième année magistère d'informatique l'École Norm. Supérieure Lyon*, 10, 1055-1064. DOI: <https://doi.org/10.1109/72.788646>
- Chen, R., Li, X. and Li, J., 2018. Object-based features for house detection from RGB high-resolution images. *Remote Sens.*, 10, 1-24. DOI: <https://doi.org/10.3390/rs10030451>
- Choraś, M., Bhanu, B., Chen, H., Champod, C., Komatsu, N., Nakano, M., Cavoukian, A., Stoianov, A., Zhou, Z.-H., Kukula, E. P., Elliott, S. J., Sarkar, S. and Liu, Z., 2009. Ensemble Learning. *Encycl. Biometrics*, 270-273.
- Dahiya, S., Garg, P. K., Jat and M. K., 2013. Object oriented approach for building extraction from high resolution satellite images. *3rd IEEE Int. Adv. Comput. Conf.*, 1300-1305. DOI: <https://doi.org/10.1109/IAdCC.2013.6514416>
- Das, M. and Ghosh, S. K., 2017. A deep-learning-based forecasting ensemble to predict missing data for remote sensing analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 10, 5228-5236. DOI: <https://doi.org/10.1109/JSTARS.2017.2760202>
- DigitalGlobe, 2014. WorldView-2 scene 053950035020_01_P001, Level Standard 2A, DigitalGlobe, Muscat, Oman.
- DigitalGlobe, 2014. WorldView-3 scene, DigitalGlobe, Jaipur, India.
- Dornaika, F., Moujahid, A., El Merabet, Y. and Ruichek, Y., 2016. Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors. *Expert Syst. Appl.*, 58, 130-142. DOI: <https://doi.org/10.1016/j.eswa.2016.03.024>
- Estornell, J., Martí, J. M., Sebastiá, T. and Mengual, J., 2013. Principal component analysis applied to remote sensing. *Model. Sci. Educ. Learn.*, 6, 83-90.
- Foody, G. M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sens. Environ.*, 114, 2271-2285. DOI: <https://doi.org/10.1016/j.rse.2010.05.003>
- Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300. DOI: <https://doi.org/10.1016/j.patrec.2005.08.011>
- Gunes, F., Wolfinger, R. and Tan, P.-Y., 2017. Stacked ensemble models for improved prediction accuracy. *Sas*, 1-19.
- Guo, Z., Shao, X., Xu, Y., Miyazaki, H., Ohira, W. and Shibasaki, R., 2016. Identification of village building via Google Earth images and supervised machine learning methods. *Remote Sens.*, 8. DOI: <https://doi.org/10.3390/rs8040271>
- He, C., Shi, P., Xie, D. and 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, 213-221. DOI: <https://doi.org/10.1080/01431161.2010.481681>
- Horning, N., 2010. Random Forests: An algorithm for image classification and generation of continuous fields data sets. *Int. Conf. Geoinformatics Spat. Infrastruct. Dev. Earth Allied Sci.*, 2010 1-6. DOI: <https://doi.org/10.5244/C.22.54>
- Ibrahim, M. A. H. Bin, Mamat, M. and June, L.W., 2014. BFGS method: A new search direction. *Sains Malaysiana*, 43, 1591-1597.
- Joelsson, S. R., Benediktsson, J. A. and Sveinsson, J. R., 2005. Random forest classifiers for hyperspectral data. *Int. Geosci. Remote Sens. Symp.*, 1, 160-163. DOI: <https://doi.org/10.1109/IGARSS.2005.1526129>
- Kupidura, P. and Jakubiak, M., 2009. The morphological filtering of the remote sensing images for the noise reduction comparing to traditional filters. *Ann. Geomatics*, 7, 63-68.
- Lari, Z. and Ebadi, H., 2007. Automatic extraction of building features from high resolution satellite images using artificial neural networks. *Proc. ISPRS Conf. Inf. Extr. from SAR Opt. Data, with Emphas. Dev. Ctries.*, CD.
- Lary, D. J., Alavi, A. H., Gandomi, A. H. and Walker, A. L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.*, 7, 3-10. DOI: <https://doi.org/10.1016/j.gsf.2015.07.003>
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S., 1996. Application of neural

- networks to modelling nonlinear relationships in ecology. *Ecol. Modell.*, 90, 39-52. DOI: [https://doi.org/10.1016/0304-3800\(95\)00142-5](https://doi.org/10.1016/0304-3800(95)00142-5)
- Lin, C. and Nevatia, R., 1998. Building detection and description from a single intensity image. *Comput. Vis. Image Underst.*, 72, 101-121. <https://doi.org/10.1006/cviu.1998.0724>
- Lindgren, J. E. and Kilston, S., 1996. Projective pan sharpening algorithm. *Proc. SPIE 2818, Multispectral Imaging for Terrestrial Applications*, 128. DOI: <https://doi.org/10.1117/12.256087>
- Martin, B. W. and Vatsavai, R. R., 2013. Image change detection via ensemble learning. *Proc. SPIE 8743, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*, 874305. DOI: <https://doi.org/10.1117/12.2017954>
- Patel, N. and Mukherjee, R., 2015. Extraction of impervious features from spectral indices using artificial neural network. *Arab. J. Geosci.*, 8, 3729-3741. DOI: <https://doi.org/10.1007/s12517-014-1492-x>
- Puttinaovarat, S. and Horkaew, P., 2017. Urban areas extraction from multi sensor data based on machine learning and data fusion. *Pattern Recognit. Image Anal.*, 27, 326-337. DOI: <https://doi.org/10.1134/S1054661816040131>
- Ramachandra, T. V., Bharath, H. A. and Sannadurgappa, D., 2012. Insights to urban dynamics through landscape spatial pattern analysis. *Journal of Applied Earth Observation and Geoinformation*, 18, 329-343.
- Ramachandra, T. V., Bharath, H. A. and Sowmyashree M. V., 2015. Monitoring urbanization and its implications in a mega city from space: Spatiotemporal patterns and its indicators. *Journal of Environmental Management.*, 148, 67-91. DOI: <https://doi.org/10.1016/j.jenvman.2014.02.015>
- Ramachandra, T. V., Bharath, H. A., Vinay, S., Kumar, U., Venugopal, K. R. and Joshi, N. V., 2016. Modelling and Visulation of urban trajectory in 4 cities of India. In Proceedings of IISc-ISRO STC 2016. January 7-8, IISc, Bangalore, India.
- Sahar, L., Candidate, P.D. and French, S.P., 2006. Using high-resolution aerial photography and neural networks to inventory properties at risk of earthquakes. *MAPPs/ASPRS 2006 Fall Conference*.
- Sinha, P., Verma, N. K. and Ayele, E., 2016. Urban Built-up Area Extraction and Change Detection of Adama Municipal Area using Time-Series Landsat Images. *Int. J. Adv. Remote Sens. GIS*, 5, 1886-1895. DOI: <https://doi.org/10.23953/cloud.ijarsg.67>
- Verticale, G. and Giacomazzi, P., 2008. Performance evaluation of a machine learning algorithm for early application identification. 2008 *Int. Multiconference Comput. Sci. Inf. Technol.*, 845-849. DOI: <https://doi.org/10.1109/IMCSIT.2008.4747340>
- Waske, B., Benediktsson, J. A., Arnason, K. and Sveinsson, J. R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. *Canadian Journal of Remote Sensing.*, 35, 106-116. DOI: <https://doi.org/10.5589/m09-018>
- Wu, Q., Wang, L. W. and Wu, J., 2012. Ensemble learning on hyperspectral remote sensing image classification. *Advanced Materials Research*, 546-547, 508-513.
- Xu, H., 2008. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.* 29, 4269-4276. DOI: <https://doi.org/10.1080/01431160802039957>
- Xu, Y., Wu, L., Xie, Z. and Chen, Z., 2018. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.*, 10. DOI: <https://doi.org/10.3390/rs10010144>
- Yuan, J., 2016. Automatic building extraction in aerial scenes using convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (11), 2793-2798. DOI: <https://doi.org/10.1109/TPAMI.2017.2750680>
