EVALUATING THE PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR URBAN LAND USE MAPPING USING VERY HIGH RESOLUTION

SHEIKH MUSTAK
April 2018

SUPERVISORS:
dr, Monika Kuffer
dr. Claudio Persello

EVALUATING THE PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR URBAN LAND USE MAPPING USING VERY HIGH RESOLUTION

SHEIKH MUSTAK
Enschede, the Netherlands, April 2018

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban planning and Management

SUPERVISORS:
dr. Monika Kuffer
dr. Claudio Persello

THESIS ASSESSMENT BOARD:
Prof. dr. K. Pfeffer (Chair)
dr. M. Netzband (External Examiner, University of Wuerzburg)

ABSTRACT

The mapping of urban land use from very high resolution (VHR) imagery is a very challenging task in remote sensing in particular in developing country like India, where land use is complex in many aspects (e.g., shape, size, orientation, etc.). In this regard, machine learning algorithms like Support Vector Machine with Radial Basis Function (SVM-RBF) and Convolutional Neural Networks (CNN) offer opportunities to improve mapping result compared to classical parametric classifiers. This study mapped the urban land use in Bengaluru city, India. The main object of this study was to evaluate the performance of different machine learning algorithms such as SVM-RBF and CNN for urban land use mapping from VHR imagery. For this purpose, different aggregation levels (beyond pixels) were employed within the machine learning algorithms using object (OBIA) and block-based (BBIA) image classification approaches. In the object-based urban land use classification, multi-resolution segmentation optimized the with the estimate scale parameter (ESP) tool was carried to obtain segments of homogeneous land uses. Regular grids were employed in BBIA. The size of the grid was selected based on literature review, and local context. In addition, several image features (i.e., spectral, textural, geometric and contextual) were extracted and aggregated from VHR imagery as well as best features, parameters, and size of training samples were explored for OBIA and BBIA-based urban land use classification using SVM-RBF. For the CNN-BBIA-based urban land use classification, best learning and regularization parameters, CNN hyperparameters and size of training samples were explored. All of the above classifications were carried out on both sampled (i.e., from the tile where training samples were taken) and unsampled domain (i.e., from the tile where training samples not taken) to assess the domain adaptability of the classifiers. For validation, different accuracy assessment indices (e.g., overall, user and producer accuracy, kappa, etc.) were measured and beyond the classical accuracy measures several other accuracy assessment indices have been used (e.g., recall, precision, F1-score Klocation, etc.). In addition, the visual quality of the classified map was compared with the referenced map (local land use map) while computational time was compared between the classifiers.

It was observed that overall accuracies of SVM-RBF-OBIA outperforms the SVM-RBF-BBIA and similarly, accuracies of CNN-BBIA outperforms the SVM-RBF-BBIA. Therefore, based on the performance evaluation it is concluded the OBIA is more relevant and robust for urban land use mapping from VHR imagery for the Indian context as compared to BBIA because urban land use is more related to the geometry (e.g., shape, size, area, etc.) of the land use. Similarly, CNN is more relevant and robust for urban land use mapping from VHR imagery for the Indian context as compared to SVM-RBF because CNN learned more complex contextual features which is essential for classifying complex land use. Therefore, this study provides a promising a starting guideline for the urban planner and local government to select appropriate machine algorithm and classification approach for efficiently mapping urban land use from VHR imagery. However, the classification accuracy in this research somehow low to implement planning policy. In this regard, the use of CNN combined with OBIA could be promising to develop a robust urban land use classification approach for VHR imagery. In addition, integration of some additional aggregation levels such parcels as blocks or additional data such as height information from a digital surface model (DSM), integrating conditional random field (CRF) with CNN might allow to improve the accuracies of urban land use classification.

Keywords: machine learning, earth observation, urban land use, image classification, transferability.

ACKNOWLEDGEMENTS

Education offers one to train the mind to build the nation and getting an education is a result of hardwork, support of family, teachers, friends, institution, and blessing of God. With heartedly, I thank God for helping me to complete my study. With immense respect, I sincerely thank my family for supporting my research. With immense pleasure, I thanks ITC, University of Twente, the Netherlands for providing me excellence scholarship for financial support. I thank my supervisors dr, Monika Kuffer and dr. Claudio Persello for providing valuable suggestions in my study. I sincerely thank all teaching and non-teaching staff members of the ITC for their co-operation in my research. I extend my gratitude to dr. Naresh Kumar Baghmar, Professor of the school of studies in geography, Pt. Ravishankar Shukla University, Raipur, for his great inspiration and heartful co-operation. I extend my thanks to all of my friends for their best wishes.

Last but not the least, finally I thank my family and God for helping me to complete my study.

TABLE OF CONTENTS

LIST OF FIGURESS

LIST OF TABLES

1. INTRODUCTION

1.1. Background and significance

More than half (54 percent) of the world's population lives in urban areas, with a projected increase to 66 percent until 2050. In total, nearly 90 percent of this increase will concentrate in Asia and Africa (United Nations, 2015), affecting mostly low and medium income countries where the majority of the urban population is concentrated in highly urbanized centres. This rapid urbanization is driven by rural-urban migration. Migrants are attracted by job opportunities and better urban facilities, but due to the absence of appropriate planning policy, a large part of this growth is concentrated in informal settlements.

This rapid urbanization is an increasing threat to the future sustainable development. For example, rapid urbanization associated with unplanned urban growth leads to high consumption of natural resources, congested and unorganized urban developments (United Nations, 2015). Unorganized urban developments lead to a complex arrangement of urban land uses (Kuffer & Barros, 2011) affecting most cities in the global south (Gevaert et al., 2017; Sandborn & Engstrom, 2016). The location, orientation, structures, and function of urban land uses such as residential, commercial and industrial areas are very complex (Sandborn & Engstrom, 2016). In addition, some of the land uses (e.g., slum, deprived area, etc.) are hidden in official planning documents (Kuffer et al., 2017; Nijman, 2008) and accuracy of overall land uses is questionable (Nijman, 2008). Besides, in the rapidly growing urban area, information quickly gets outdated (Wentz et al., 2014). In this regard, accurate, consistent and timely information on city growth is required to support policy development towards sustainable development and prioritise policy on equitable access for present and future needs (United Nations, 2015; Wentz et al., 2014).

Mapping of urban land use from satellite imagery is challenging because of the absence of an appropriate classification method (Sandborn & Engstrom, 2016; Tewkesbury et al., 2015; Wieland & Pittore, 2014). Very high resolution imagery can acquire textural, spectral and colour characteristics of land use which could be used in general land use classification (Sandborn & Engstrom, 2016). However, the urban landscape has heterogeneous spatial patterns and complex functional characteristics, which are very hard to distinguish in discrete land use classes using only spectral information. In complex urban areas, each land use has a distinct spectral response, texture, geometry, orientation, spatial arrangement and a functional characteristic such as transportation network, residential buildings, etc. (Sandborn & Engstrom, 2016; Wieland & Pittore, 2014).

In classical urban land use mapping, spectral information is commonly used (Liao et al., 2017; Mboga, 2017; Tang et al., 2012; Wentz et al., 2008). Recent studies showed that adding additional geometrical (shape, size, area etc.) and contextual information (object-level, street block-level, and parcel-level attributes, etc.) is advantageous to improve classification accuracies (Cockx, Van de Voorde, & Canters, 2014; Herold, Liu, & Clark, 2003; Kuffer et al., 2017; Yanchen et al., 2014). In this regard, VHR imagery are advantageous to extract geometric and contextual features for detailed urban land use classification (Hu & Wang, 2013; Li et al., 2016; Ma et al., 2015; Wu et al., 2009; Yanchen et al., 2014; Yang et al., 2010; Zhang et al., 2017). Recently, VHR are widely using for mapping detailed urban land use (e.g., Li, et al., 2016), single urban land use (e.g. slums) (e.g., Naorem et al., 2016; Gevaert et al., 2017; Kohli, 2015; Kuffer et al., 2017; Mboga, 2017; Pratomo, 2016), road detection (e.g., Sameen & Pradhan, 2016) and building footprint extraction (e.g., Gokon et al., 2015; Huang & Zhang, 2012).

Pixel-based image analysis (PBIA) is a very common and widely used image classification approach which has several limitations like 'salt and pepper effect' (noise) and spectral confusion. In addition, traditional

PBIA like maximum likelihood classifier (MLC) also has several limitations such as mixed pixel classification, inability to integrate adequate contextual information and as a consequence, such classifier is unable to solve complex problems of urban land use classification (Lu & Weng, 2007). In the object-based image analysis (OBIA), pixels are grouped into the object-level in which different spectral, textural, contextual and geometric features are extracted for classification. This approach is very convenient for extracting features related to shape, size and texture observed in VHR imageries (Herold et al., 2003; Kohli, 2015; Kuffer et al., 2017; Li et al., 2016; Ma et al., 2015).

Another important image classification approach is block-based image analysis (BBIA) in which the entire image is split into blocks (e.g., regular grid, street block, etc.). After that, image features are extracted at block-level. Such extracted features at block-level can be used for urban land use classification (Herold et al., 2003; Sandborn & Engstrom, 2016). Thus, a block-based image classification approach extracts aggregated contextual information (form, shape, pattern) and is, therefore, more relevant for urban land use mapping (Herold et al., 2003; Sandborn & Engstrom, 2016; Silván-Cárdenas, Almazán-González, & Couturier, 2014). This is because, land use cannot be linked with pixels for the reason that a land use (zone) is the aggregation of several individual land cover objects (Herold et al., 2003).

Machine learning techniques are widely used in computer science, medical science, gaming technology for data mining, pattern recognition and image classifications (Persello & Bruzzone, 2014; Weiss, Khoshgoftaar, & Wang, 2016). Recently, machine learning is a widely used technique in geo-information science and earth observation for different application including urban land use mapping (Berger et al., 2013; Wieland & Pittore, 2014), change detection analysis (Tewkesbury et al., 2015), road networks (Sameen & Pradhan, 2016) and building footprints extraction (Gokon et al., 2015; Huang & Zhang, 2012), slum delineations (Gevaert et al., 2017) and urban village mapping (Liu et al., 2017) due to its smart, fast and cutting-edge ccomputational performance (Persello & Bruzzone, 2014; Weiss, Khoshgoftaar, & Wang, 2016; Wieland & Pittore, 2014). In addition, the performance of machine learning algorithms on VHR imagery is promising due to their high capability of data integration, automatic learning of training samples, non-linear computation, customized algorithms and handling of wide-scale image analysis elements which are very essential for solving complex problems in the urban land use mapping (Persello & Bruzzone, 2014; Weiss, Khoshgoftaar, & Wang, 2016).

In this regard, supervised machine learning algorithms such as support vector machine (SVM), decision tree (DT), random forest (RF), K-Nearest Neighbours are increasingly used for urban land use mapping (Berger et al., 2013; Wieland & Pittore, 2014). The Convolution Neural Network (CNN) is an advanced deep learning algorithm which is recently used by several researchers for urban land use classification by taking advantages of its self-extracting capability of image features (Bergado, Persello, & Gevaert, 2016; Lee & Kwon, 2016; Mboga, 2017). Thus, to overcome limitations of PBIA and traditional image classifiers, OBIA and BBIA are highly encouraged to be integrated with machine learning classifiers to improve the overall classification accuracy (Chuang & Shiu, 2016; Tewkesbury et al., 2015; Wieland & Pittore, 2014). In image classification, both SVM-RBF and CNN are robust machine learning algorithms (Mboga, 2017; Stavrakoudis et al., 2014; Tang et al., 2012) which have rarely been combined with OBIA and BBIA for urban land use classification of Indian cities. In this regard, SVM-RBF and CNN algorithms have been selected to test their performances on BBIA, and OBIA for classifying urban land uses in Bengaluru city, India. This study is carried out based on an experimental research design in which the performance of selected machine learning algorithms is evaluated using reference datasets.

1.2. Research problem

Based on the literature review, the research problem in this study has been explained below (see figure 1): Firstly, there is insufficient knowledge about the best image features for urban land use classification.

Secondly, to classify complex urban land use it necessities to use advanced and robust machine learning algorithms like SVM (e.g., Silván-Cárdenas et al., 2014) and CNN (e.g., Hu et al., 2015) which were rarely used for the Indian context. Thirdly, the classification of urban land use is more appropriate with OBIA and BBIA because the concept of land use related to the larger area instead of the pixel (Herold et al., 2003). In general, OBIA (e.g., Man, Dong, & Guo, 2015) and BBIA (e.g., Sandborn & Engstrom, 2016) have not been much used for urban land use classification, and similarly, this also has not been previously explored for Indian cities. Finally, land use classification in India is commonly done by visual image interpretation with field survey in several city planning departments which is slow, time consuming, costly, and information gets quickly outdated. This study will help to resolve the above research gap by employing advanced image analysis techniques.

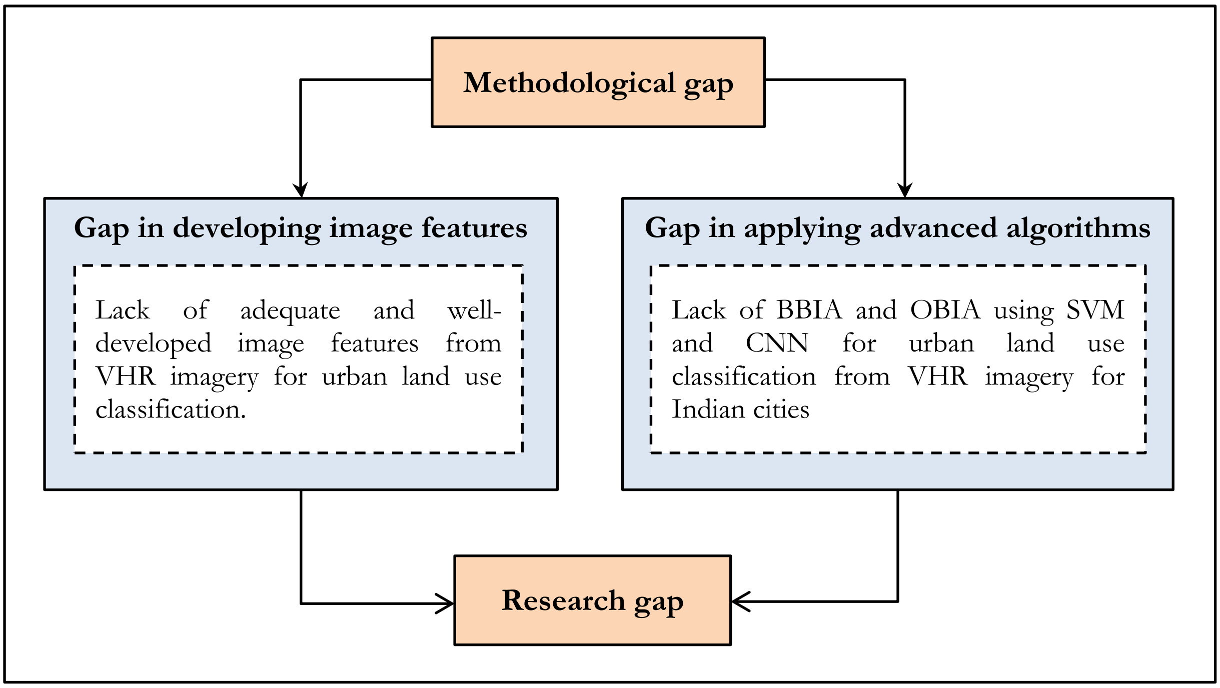


Figure 1. Shows research gaps.

1.3. Research objectives

The main objective of this study is to evaluate the performance of different machine learning algorithms for urban land use mapping.

1.3.1. Specific objectives

The specific objectives are outlined to carry out the proposed research which as follows-

1. To select suitable image features for urban land use mapping.
2. To map urban land uses using SVM and CNN in OBIA and BBIA.
3. To evaluate the performance of SVM and CNN in OBIA and BBIA for urban land use classification.

1.4. Research questions

Specific objective 1

1. What types of image features are extracted from VHR imagery using standard feature extraction methods based on recent literatures?

2. What is the standard feature selection method used for selecting best features based on recent literatures?
3. What are best the image features used to map urban land use using standard feature selection method?

Specific objective 2

1. What types of urban land uses are relevant based on national and local land use classification schemes and available literatures?
2. What are best parameters of SVM and CNN for classifying urban land uses employing OBIA and BBIA?
3. What are the classification accuracies and time elapses executing a SVM in BBIA and OBIA employing the best parameters and image features?
4. What are the classification accuracies and time elapses executing a CNN in BBIA employing the best parameters?

Specific objective 3

1. What is the best strategy to measure the accuracy of SVM and CNN for urban land use classification?
2. What is the performance of SVM and CNN for urban land use classification?

1.5. Conceptual framework

This research is conceptualized in figure 2. The features extraction from the VHR imagery is a very primary concept in image classification.

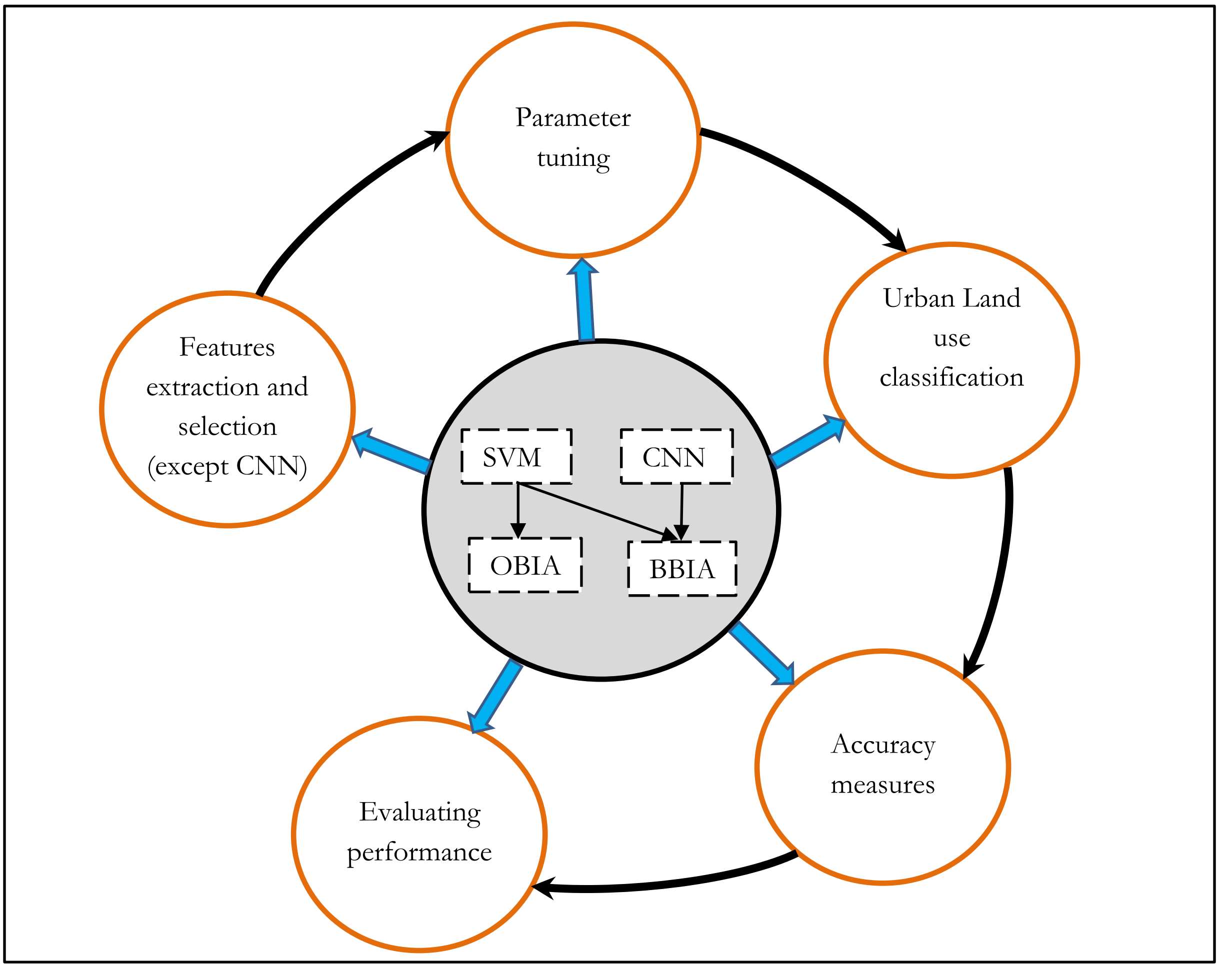


Figure 2. Shows conceptual framework.

The concept of features extraction relates to the extraction of different spectral (e.g., spectral bands, NDVI, etc.), textural (e.g., GLCM, LBP, etc.), geometric (e.g., shape index, roundness, etc.) and contextual features (e.g., spatial matrices). In addition, another important concept is features selection which is used to select the best features that reduced the effects of Hughes phenomena and enhance the classification accuracy and computational performance (Camps-valls, Mooij, & Schölkopf, 2010; Persello & Bruzzone, 2016; Damodaran, Courty, & Lefevre, 2017; Niazmardi, Safari, & Homayouni, 2017). Thus, next to the features selection, the parameter tuning is one of the very important concept in image classification which is used to select the best parameter to reduce the risk of overfitting of the classifiers and improve the classification accuracy. The concept of image classification relates to the approach (e.g., OBIA and BBIA) that is used in urban land cover/use classification. This incorporates the concept of features extraction, features selection, parameter tuning, and algorithms (e.g., SVM, CNN) used. Finally, accuracy measures and performance evaluation in image classification are used to validate and assesses the ability of the classifier to map urban land use.

1.6. Thesis structure

Chapter 1- provides the background and justification for selecting the research topic, research gap identification, research objectives, research question and conceptual framework.

Chapter 2- provides a detailed literature review about pixel-based, object-based and block-based urban land use classification. This chapter also gives an overview of different types of machine learning algorithms, image features, feature extraction, and selection methods, parameter tuning methods, different types of urban land use classification scheme, urban land use pattern in Indian cities, etc.

Chapter 3- introduces the study area, the dataset used, tools and software used. In this chapter, the physical and demographic status of the study area is discussed and similarly explained about the remote sensing and other referenced datasets. This chapter, provides a detailed discussion about the methods of pre-processing, block generation, image segmentation, features extraction, features selection, parameter tuning, architectures of SVM and CNN, land use classification, accuracy measures and performance assessment.

Chapter 4- covers results and discussion such as extracted features, selected best features, best parameters, urban land cover and land use classification based on SVM-RBF-based OBIA and BBIA and CNN-based BBIA, etc. This chapter also covers the measuring accuracies of SVM-RBF-based OBIA and BBIA and CNN-based BBIA for urban land use and cover classification.

Chapter 5- provides the conclusions and recommendations. In this section, synthesizing the research results by addressing the research objectives, research questions, limitations and drawing final recommendations for future research.

2. LITERATURE REVIEW

The literature review has been carried on different approaches, methods, and contexts related to urban land use mapping using VHR imagery.

2.1. Pixel-based image classification approach

The PBIA is the standard and widely used approach in the field of remote sensing (Lu & Weng, 2007). The PBIA is linked with the two concepts such as type one PBIA in which both input features and classification output are purely pixel-based (e.g., spectral band, NDVI features with SVM.). For example, Tang et al.(2012) were used spectral features (e.g., spectral band and others spectral features from Landsat TM) to evaluate the performance of the MLC, classification tree, Random forest, Bagging and SVM for urban land use classification of New Orleans and Baton Rouge, USA. The authors argued that SVM is the most robust classifier as compared to the others classifiers. Secondly, type two PBIA in which input features are computed from the neighbourhood (e.g., patch-based or window-based) of a single pixel and classification output is pixel-based (e.g., CNN; GLCM, local binary pattern, morphological profile features with SVM) which highly linked with the block-based image classification approach. For example, Liao et al. (2017) were compared different image features using SVM for urban land use classification of the city of Ghent Belgium; Pavia, Italy; and Houston, USA. The authors argued that SVM with morphological features (e.g., type two PBIA) outperform the SVM with spectral bands (e.g., type one PBIA). In addition, Mboga (2017) was also used similar PBIA approach to compare the performance of SVM (using GLCM and CNN features from VHR IKONOS, hyperspectral imageries) with CNN for classifying urban land use like the formal and informal settlement of Dar es Salaam, Tanzania. The author argued that CNN outperform the SVM because of the high computational performance of complex contextual image features. Therefore, it is concluded that type two PBIA with an advanced classifier (e.g., CNN) outperform the type one PBIA as well as type two PBIA with a commonly used classifier (e.g., SVM, random forest, etc.) for urban land use classification.

In addition, some of the studies employed the mixed PBIA approach in which input features computed pixel-wise and neighbourhood-wise, but classification output is pixel-wise. For example, Wentz et al. (2008) used both spectral (e.g., spectral band, NDVI, etc. from ASTER imagery) and contextual features (e.g., GLCM from ASTER imagery) for classifying the urban land use of Delhi city, India using expert decision tree classifier. However, using PBIA, several urban land use classes such as residential, commercial, industrial, transportation etc. were classified in different studies outside India (e.g., Liao et al., 2017; Mboga, 2017; Tang et al., 2012) but very limited study was carried out in India (e.g., Wentz et al., 2008). In addition, for an Indian city (Delhi), Wentz et al. (2008) were unable to classify important urban land uses such as residential, commercial, etc. perhaps due to the limitations of the coarse resolution imagery (Aster) and the traditional image classifier. The above studies reveal that the integration of textural and contextual features along with spectral features from VHR imagery in PBIA is very useful to classify detailed urban land use. In addition, the above literatures also claimed that the SVM and CNN are robust classifiers. CNN is a new and advanced classifier recently being applied for urban land use classification.

2.2. Object-based image classification approach

The OBIA is a widely used image classification approach employed in several studies for urban land use mapping (e.g., Kuffer et al., 2017; Ma et al., 2015; Man, Dong, & Guo, 2015; Yanchen et al., 2014). In

OBIA, both input features and classification output are based on object-level instead of pixel. The advantages of OBIA is that it included the geometric features (e.g., shape, size, etc.) along with others spectral (e.g., NDVI, nDSM, etc.), textural (e.g., GLCM, etc.) and contextual features (e.g., spatial metrics) at object-level. For example, Ma et al. (2015) used geometric (e.g., regularity) and contextual features (e.g., lacunarity) while Yanchen et al.(2014) included spectral (e.g., GLCM), contextual and geometric (e.g., asymmetry, border length, IHS transformation etc.) for urban land use classification. In addition, few researchers (e.g., Kuffer et al., 2017) integrated PBIA with OBIA to improve the classification of urban land use. This integration was used to include some important contextual features (e.g., spatial metrics from land cover) into OBIA for urban land use classification (Kuffer et al., 2017).

Thus, Yanchen et al.(2014) were compared the performance of PBIA (using GLCM features) with OBIA using SVM for mapping urban land use of near Western 3rd Ring Road of Beijing, China. The authors argued that SVM with OBIA outperform the SVM with PBIA for urban land use classification. In OBIA, several VHR imageries such as VHR Aerial color images (e.g., Ma et al., 2015), VHR WorldView-2 (e.g., Kuffer et al., 2017), VHR hyperspectral (e.g., Man, Dong, & Guo, 2015) and LiDAR (e.g., Ma et al., 2015; Man, Dong, & Guo, 2015) were widely used for urban land use classification. The above study shows that OBIA was commonly used for the extraction of important urban land use such as residential, commercial, residential, transportation, etc. of the other than Indian cities (e.g., China, USA, etc.). Therefore, it is concluded, integration of PBIA with OBIA and is an additional advantage for improving the urban land use classification accuracy using VHR imagery. Furthermore, it also concluded that the use of robust classifier such as SVM provides added advantages on OBIA for urban land use classification.

2.3. Block-based image analysis approach

The BBIA is an important image classification approach widely used for several studies for correlating or extracting urban land use from VHR imagery (e.g., Duque, Patino, & Betancourt, 2017; Sandborn & Engstrom, 2016; Silván-Cárdenas et al.,2014). In BBIA, both input features and classification results are based on block-level (e.g., regular grid, parcel, etc.). However, some other PBIA approach such as CNN strongly linked with BBIA because CNN aggregated more abstract features at patch-level (CS231n, 2018; Bergado, 2016; Mboga, 2017). In addition, in BBIA some of the studies used regular grid (e.g., Duque, Patino, & Betancourt, 2017; Herold et al. 2003; Sandborn & Engstrom, 2016) while others used parcel (e.g., Hu & Wang, 2013; Silván-Cárdenas et al.,2014) or road network grid (e.g., Li et al.,2016) for correlating or classifying urban land use. This block was either downloaded from the Openstreet map (e.g., Li et al.,2016) or prepared by manual digitization (e.,g., Hu & Wang, 2013) or automatically generated from the software (e.g., Duque, Patino, & Betancourt, 2017). Similar to the OBIA, the BBIA also included spectral (e.g., NDVI, nDSM etc.), textural (e.g., GLCM, LBP etc.), contextual (e.g., spatial metrics) and geometric features (e.g., shape, compactness etc.) (except geometric features for regular grid) for urban land use classification as observed in above studies.

It was observed that some of the studied were integrated PBIA (e.g., Cockx, Van de Voorde, & Canters, 2014; Hu & Wang, 2013) or OBIA (e.g., ., Li et al., 2016) with BBIA for improving urban land use classification accuracy. This integration was done to extract other contextual features such as spatial metrics (e.g., Herold et al. 2003), coverage ratio (e.g., Li et al., 2016), etc. from land cover and such features were used for BBIA-based urban land use classification. In BBIA, one of the very important facts is the this also can include demographic and socio-economic information of land use such as population density, tax information, etc. along with others features while parcel is conder as the block (Wu et al., 2007). However, it was also observed that some of the studied compared classifiers such as SVM with MLC (e.g., Silván-Cárdenas et al.,2014) and SVM with CNN (e.g., Mboga, 2017) for BBIA-based urban land use from VHR imagery (e.g., Quick Bird images, LiDAR, etc.). Based on the performance of these classifiers the authors concluded that SVM outperforms the MLC while CNN outperforms the SVM. In addition, from the above studies, it was also observed that BBIA was used widely used for classifying important urban land use such as residential, commercial, industrial, transportation, etc. of the cities outside India (e.g., China, Africa, USA, etc.). Therefore, from the above literatures, it is concluded that CNN with BBIA is robust classification approach as compared to SVM with BBIA which was never used for detailed urban land use classification for Indian cities from VHR imagery. Furthermore, it is also concluded that integration of PBIA or OBIA with BBIA is an added advantage for urban land use classification from VHR imagery.

2.4. Support vector machine with radial basis function

The Support Vector Machine with Radial Basis Function (SVM-RBF) is a robust classical machine learning algorithm for non-parametric as well as non-linear image classification problems (Bruzzone & Persello, 2009; Gevaert et al., 2016). The SVM is a binary classifier (0, 1) which used to solve the linear (see figure 3(A)) and non-linear classification problem (see figure 3(B)). In the non-linear SVM, non-linear mapping function $\varphi(x_i)$ (see equation 1-4) was employed to separate the two classes (y_i ∈ +1, -1) (see equation 4) corresponding to the n set of training samples (x_i) from the hyperplane (H) based on margin maximization (M) of primal quadratic optimization problem (see equation 3) which explained in equation 1-4 and figure 3 (Bruzzone & Persello, 2009; Gevaert et al., 2016; Mourão-Miranda et al., 2011).

$$y_i f(x) = w.\varphi(x_i) + b \tag{1}$$

$$H: y_i f(\varphi(x_i) = 0 \tag{2}$$

$$\min_{wb\xi} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \right\} \tag{3}$$

$$\begin{aligned} subject\ to:\ & y_i.(w.\varphi(x_i) + b \geq 1 - \xi_i, i = 1, \ldots\ldots, n \\ & \xi_i > 0, i = 1, \ldots\ldots, n \end{aligned}$$

$$\begin{aligned} Class\ seperation\ subject\ to:\ & y_i f(\varphi(x_i) > 0 \Rightarrow y_i = +1 \\ and\ \ & y_i f(\varphi(x_i) < 0 \Rightarrow y_i = -1 \end{aligned} \tag{4}$$

Where, where, w=weight vector, b=bias, C=regularization parameter corresponds to the cost of the wrong classification.

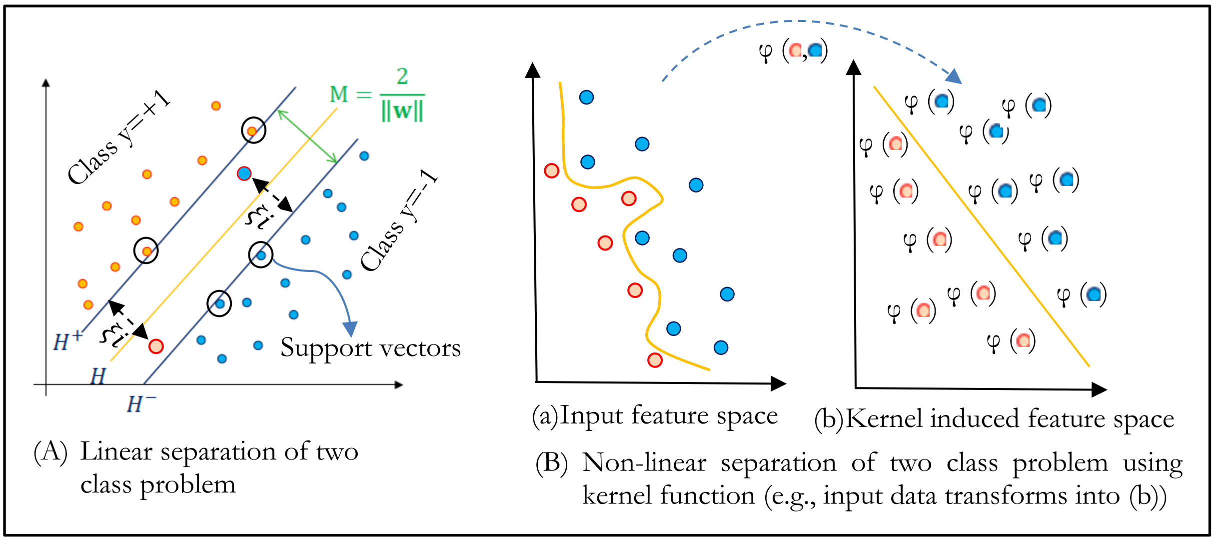


Figure 3. Shows SVM architecture for linear and non-linear class separation problem (Bruzzone & Persello, 2009).

The dual problems of quadratic optimization employed by Lagrange multipliers to solve the primal quadratic optimization problem (see equation 6). In addition, non-linear mapping function (φ(x_i). φ(x_j)) is replaced by the kernel function of Gaussian Radial Basis Function (K_{RBF} (x_i, x_j) = φ(x_i). φ(x_j)) (equation 5) to developed SVM-RBF which explained in equation 5-6 (Gevaert et al., 2016).

$$K_{RBF}(X_i, X_j) = \exp(-\frac{\|X_i - X_j\|^2}{2\sigma^2} \tag{5}$$

$$\max_{a}\left\{\sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i\, a_j K_{RBF}(x_i, x_j)\right\} \tag{6}$$

$$subject\ to: \sum_{i=1}^{n} y_i\, a_i = 0, 0\ \leq a_i \leq C, i = 1, \dots, n$$

Where, a_i= Lagrange multipliers, σ= bandwidth in Gaussian function is determined by the median distance of training samples. The architecture of SVM for a two class problem is solved by the common rule of one-against-other (OAO) while a multiclass problem is solved by one-against-all (OAA) (Richards & Jia, 2006). As mentioned in the equation, the SVM-RBF classifier is controlled by two classification parameters, i.e., cost (C) and gamma (σ) (Suykens,2001). The cost parameter is used to control the slack variables (noise) and outliers while the gamma parameter used to control the width of RBF kernel of the variables from the decision plane (Gevaert et al., 2016; Mourão-Miranda et al., 2011). Both the cost and gamma parameters control the classification accuracy (Persello & Bruzzone, 2014; Suykens, 2001). The optimum cost and gamma parameters reduce the risk of overfitting and improve classification accuracy which could be obtained by the parameter tuning using grid search hold-out k-fold cross-validation function (Persello & Bruzzone, 2014).

2.5. Artificial neural network

The Artificial Neural Network (ANN) is one of the very advanced computing algorithm in computer science which derived its name from the human biological neuron(Atkinson & Tatnall, 1997; Madani, 2008). In human biological nervous systems, different neurons are connected to the others neuron through the synapses which is a biological neural network (BNN). In a BNN, neuron (consist of cell body, axon, and dendrite) collects input information (e.g., taste) from the biological detectors (e.g., tongue) and passing this information to the reflector (e.g., brain) through interconnected neurons to take the final decision (e.g., types of taste) in feedforward and feedback response process (Atkinson & Tatnall, 1997). In this process, input information (x_1, $x_{2..,...,}x_n$) is collected through dendrite and is transformed into electrochemical signal through activation in the cell body, and this activated signal is transmitted from one neuron to another neuron through an axon.

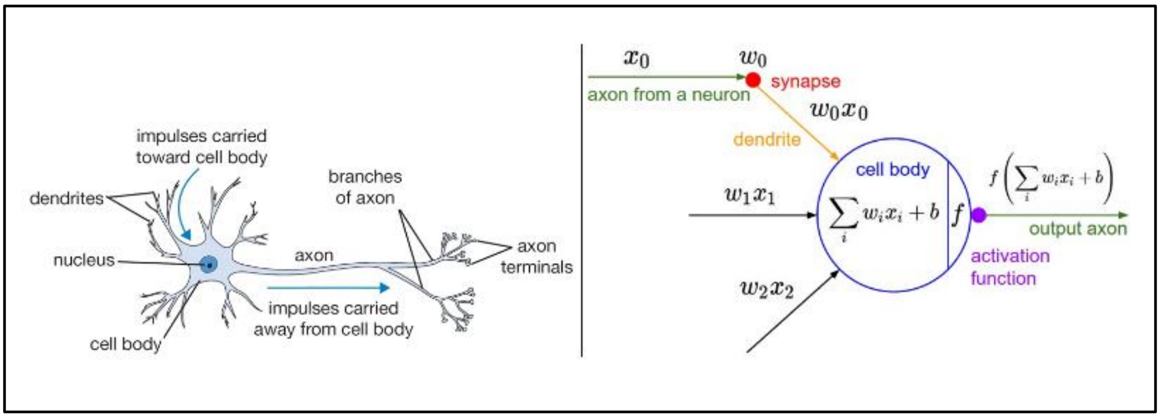


Figure 4. Shows structure and function of ANN similar to biological BNN adopted from (CS231n, 2018).

The activation of the neuron performs with the weight (w) obtained from the synapses (connectors of two neurons) and unit bias (b) which explained in figure 4 (CS231n, 2018; Mboga, 2017). Thus, later scientist transformed the principle of BNN into ANN to solve several complex problems using advanced mathematical and statistical computations. Initially, the ANN was used for pattern recognition (Kim et al., 2012; Kim, 2010; Madani, 2008) but recently it is increasingly used in remote sensing for land use/cover classification (e.g., Bergado, 2016; Civco, 1993; Paoletti et al., 2017; Yuan, Van Der Wiele, & Khorram, 2009). The neuron in the ANN is termed as perceptron and based on the arrangement of layers in the network is called a single layer perceptron (SLP) or multilayer perceptron (MLP) which explained below.

2.5.1. Single layer perceptron

The single layer perceptron (SLP) is a very simple form of an ANN which was initially designed by McCulloch-Pitts (1943) to solve the linear separable problem of two classes from the hyperplane(P) in a n-dimensional space based on the sign activation function (sgn) (Zhang & Zhang, 1999). The SLP consists of a single neuron with input (x_i) and output (y). The weighted sum of the input is calculated using input (x_i), weight factor (w_i) and threshold (e.g., bias) using the equation 7. The weighted sum of the input is activated in feedforward process using sign activation function followed by either OR, AND or NOT Boolean logic to solve the simple linear classification problems which are explained in equation 7-8 and figure 5 (Degeratu, Schiopu, & Degeratu, 2001). The equation 7 and 8 explained that when a weighted sum of the input is less than the threshold (x_p) of the neuron, the output will be -1 otherwise 1.

$$y = sgn\left(v = \sum_{i}^{n} w_i . x_i - x_p\right) \tag{7}$$

$$sgn(v) = \begin{cases} 1 & v > 0 \\ 0 & v = 0 \\ -1 & v < 0 \end{cases} \tag{8}$$

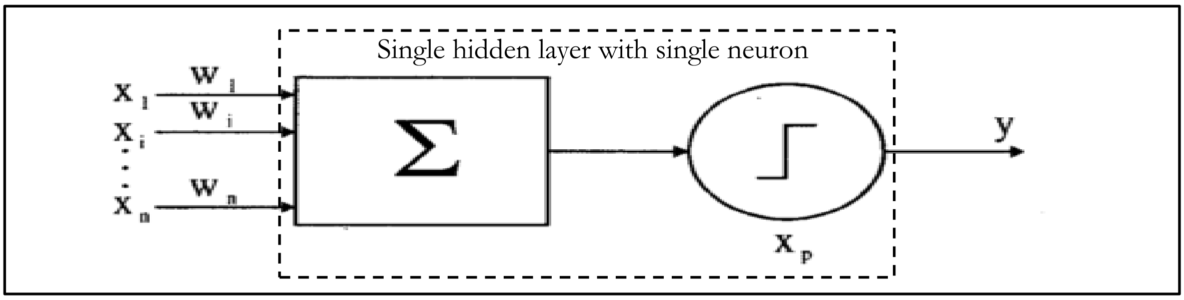


Figure 5. Shows SLP scheme of McCulloch-Pitts (1943) adopted and modified from (Degeratu et al., 2001).

The limitation of SLP is that it is unable to provide a non-linear solution of a multiclass class problem because SLP can only adopt the AND, OR and NOT Boolean logic for linear separable of two classes (Degeratu et al., 2001). The non-linear solution for complex multiclass classification problem required more neurons which could be solved with XOR Boolean logic (Kim & Choi, 1997). Later, multiclass classification problems were introduced in ANN through advanced activation function with feedforward and backpropagation algorithms which explained in MLP.

2.5.2. Multilayer perceptron

The multilayer perceptron (MLP) is an advanced ANN over SLP-ANN which is configured with input layer(x_i), hidden layers and output layer (y_i) (Lin, 2011) (see figure 6). In MLP, the result is obtained through feedforward computation of the weighted sum of learned input (x_i) with weight(w_i) and bias (b_i) using non-linear activation function (e.g., sigmoid, tanh) followed by XOR Boolean logic (Lin, 2011). In

MLP, each neuron of the input, hidden or output layers are fully connected with all neurons of the preceding or succeeding layers (Lin, 2011) (see figure 6).

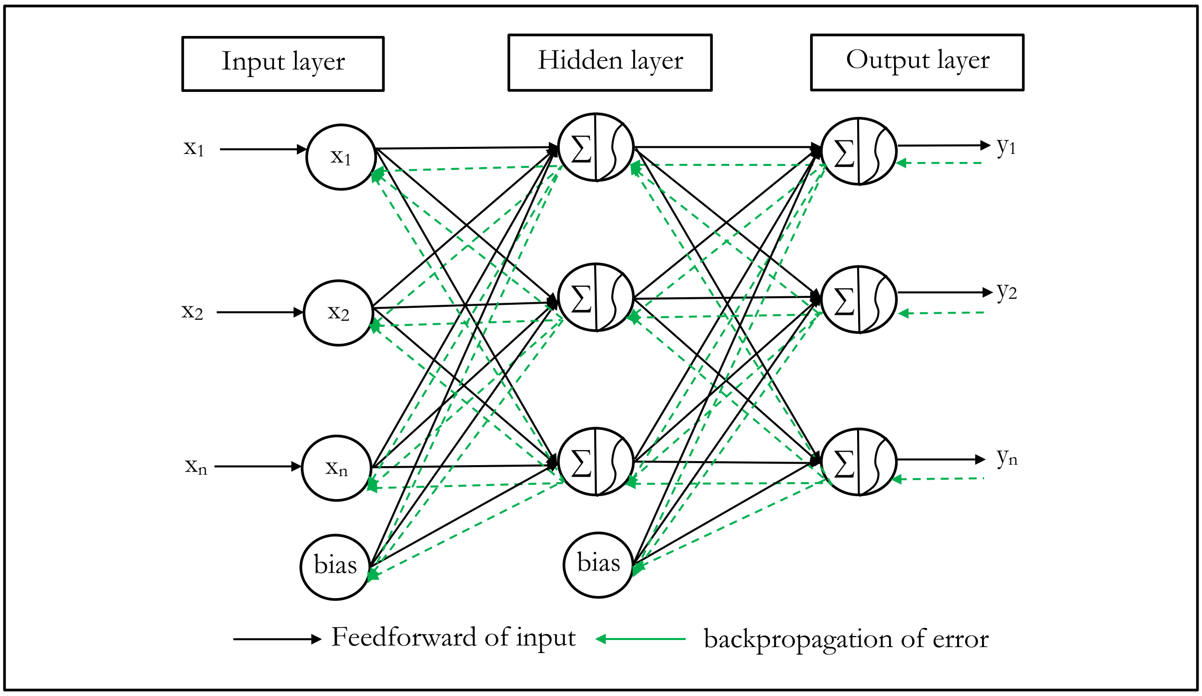


Figure 6. Shows ANN-MLP architecture with input layers (three neurons), one hidden layer (three neurons) and output layer (three neuron).

In the MLP learning process, a weighted sum of each neuron is computed using the input (x_i), weight (w_i) and unit bias (b_i) which is explained in the equation 9 (CS231n, 2018).

$$y_i = \sum_{i}^{n} w_i . x_i + b_i \qquad (9)$$

The weighted sum of the neuron (y_i) is activated using most commonly used non-linear activation functions such as sigmoid and hyperbolic tangent function (tanh) to get the output (o_i) which explained in the equation 10 and 11 (Lin, 2011; Mboga, 2017).

$$o_i = tanh(y_i) \qquad (10)$$

$$o_i = \frac{1}{1 + e^{-y_i}} \qquad (11)$$

The output of the neuron (o_i) in layer (l_i) is fully connected with the succeeding neuron in the next layer(l_{i+1}) as an input. Thus, the network feeds the input to the neuron of the next layer to get the final output, which is called as feedforward process in the network perceptron phase (e.g., perceptron learning rule). The perceptron learning rule with activation function is termed as the generalization phase of MLP. However, while employing function and architecture of MLP in deep learning (e.g., CNN), two different advanced activation functions (e.g., RELU, SoftMax) were applied for two different types hidden layers. Firstly, rectified liner unit (RELU) activation function was applied in the hidden layer (e.g., convolutional layers) instead of sigmoid function because RELU is fast and robust as compared to the other activation functions (CS231n, 2018). Thus, activation output (o_i) of the weighted sum of the neuron (y_i) in the hidden layer was computed using RELU activation function which is explained in the equation 12 (CS231n, 2018).

$$o_i = Relu(y_i) \tag{12}$$
$$where, \qquad Relu = f(y_i) = \max(0, y_i)$$

Secondary, SoftMax activation function was applied in the output layer of the last hidden layer (o_i) to get the final output probability (e.g., between 0-1) (e.g., posterior probability) of the different class label (o_j) which explained in the equation 13 (Mboga, 2017).

$$p(o_j|o_i) = \frac{\exp(o_j)}{\sum_{i=1}^{n} exp(o_i)} \tag{13}$$

Where, n=un-normalized sample of n number of class, i and j. In the generalization phase of MLP, network measures the learning error ($E_n(w)$) from the *n* training samples using the equation 14 which is called as the objective loss function (e.g., cross-entropy or negative log likelihood) (Bergado, 2016).

$$E_n(w) = -\sum_i t_i \log o_i \tag{14}$$

Where, t_i= vector encoding (0 or 1) of true label vector corresponding to the i^{th} class and o_i=output probability label vector of the i^{th} class. The error measured using objective loss function is automatically minimized by updating the weight of the preceding neurons using delta rule (Δ) which is explained in equation 15 (Bergado, 2016).

$$\Delta w(\tau) = -\epsilon(\tau)\frac{\partial E(\tau)}{\partial w(\tau)} + a\Delta w(\tau - 1) \tag{15}$$

Where, w=weight, $\epsilon(\tau)$= learning rate, τ=epoch, a=momentum and $\partial E(\tau)/\partial w(\tau)$ =gradient. The delta rule is called as the backpropagation algorithm with stochastic gradient descent (SGD). This phase of the MLP is termed as the optimization phase. The optimization phase is controlled by some initialization parameter of the network such as learning rate, momentum, and epoch as explained in equation 15. The learning rate (e.g., 1/10, 1/100, 1/1000, etc.) is the fraction of learning which explained how precisely updates the derivates of the SGD to obtain minimum loss function in each epoch. In addition, learning rate also controls the speed of learning while momentum (0-1) is used to accelerate the learning in each epoch (Mboga, 2017). The epoch is the number of iteration of the network. In addition, overfitting of the network (resulted while increase depth of the network) can be managed by applying regularization parameters such as weight decay (λ) in the L2 regularization norm and dropout (e.g., randomly dropping out percent of neuron to minimize co-adaptation between neuron) (Srivastava et al., 2014; Bergado, 2016). The L2 regularization norm for the new loss function J(w) is explained in the equation 16 (Bergado, 2016).

$$J(w) = E(w) + \lambda \,\|w\|^2 \tag{16}$$

However, cross-validation is one of the options to select optimum initialization and regularization parameters to minimize the loss function in network optimization process.

2.5.3. Convolutional neural networks

In this study, the selected ANN is the convolutional neural networks (CNN) because of its advanced pattern recognition ability. The convolutional neural networks (CNN) employed function (e.g., activation, feedforward, and backpropagation, etc.) and architecture of the MLP (e.g., hidden layers) in the networks. The difference between MLP and CNN is observed in terms of architecture and depth of the network which is explained in MLP (see section 2.5.3) and in this section (Bergado, 2016). In CNN, two types of hidden layers are used. Firstly, type one hidden layer is convolutional layer which is the sparsely connected neuron, and secondly, type two hidden layer is the fully-connected layers which are similar to the fully-connected neuron of the MLP. The CNN is used for the grid-based data structure to solve non-linear multiclass classification problems (Goodfellow et al., 2016). Consequently, CNN was used for many grid-based multiclass classification problems such as land cover/use classification from satellite imagery during past several decades (e.g., Bergado, 2016; Hu et al., 2015; Lee & Kwon, 2016; Mboga, 2017; Paoletti et al., 2017).

In addition, the most commonly used architecture of CNN (e.g., in MatConNet library) was developed with input layer, convolutional layers (conv), activation layer, pooling layers, fully-connected layers (FC) and output layer (e.g., INPUT-CONV-RELU-POOLING-FC-OUTPUT) (CS231n, 2018; Bergado, 2016) see figure 7. The input layer is a 3-D volume (length, width, and depth) which consists of *n* number of 2-D image patches. As an example, eight spectral of bands of 29 X 29 image patches (29X29X8) are used as the input for the first convolution layer in this study. Similarly, the structure of the convolution layer is the 3-D volume (length, height, and depth) which includes *n* number of 2-D convolving filters (e.g., 5X5X8) that produced the equivalent size of output volume. The spatial size of the output volume of convolution layer ($Conv_{out}$) can be estimated as the function of input size (I), filter (F), stride and zero padding (P) using equation 17 and the output must be an integer (CS231n, 2018).

$$Conv_{output} = \left(\frac{I_{size} - F_{size} + 2P}{S}\right) + 1 \tag{17}$$

However, the depth of the output volume (e.g., number of activation maps) of the convolution layer ($Conv_{out}$) is user-defined. In addition, the zero padding can be determined in relation to the filter size using the equation 18 (CS231n, 2018).

$$Zero\ Padding = \left(\frac{F_{size} - 1}{2}\right) \tag{18}$$

The stride is the step of sliding filter over the spatial input while zero padding is the zero around the border of the spatial input which is used for preserving the spatial characteristic of the object (CS231n, 2018). The principle of CNN is the sparse connectivity, parameter sharing and equivariant representations (Goodfellow et al., 2016). In terms of sparse connectivity, the output unit (e.g., neuron) of the convolution layers (s) is sparsely connected with the unit of input layer (x) by the function of convolving kernel (moving filter) where the kernel size is less than the input size (Goodfellow et al., 2016).

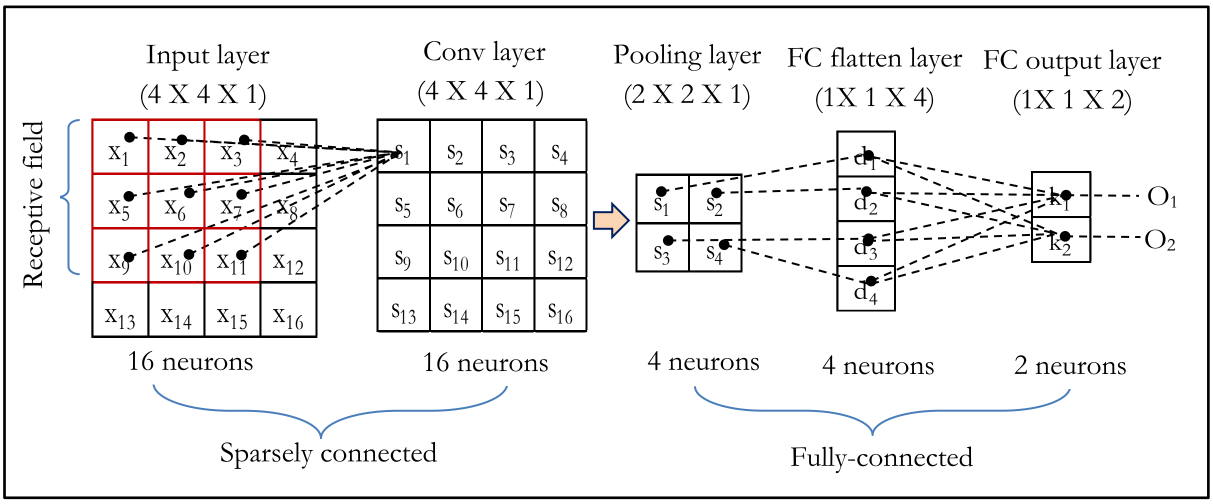


Figure 7. Shows an example of sparsely and full connected neurons in the CNN.

In case of the FC layer, the output neuron is fully-connected with all preceding input neurons as mentioned in the MLP (see figure 7). The convolving kernel (e.g., 3X3 local neighbourhood) in input layer (e.g., x) is the receptive field of the neuron of the convolutional layer (e.g., s_1) (see figure7). As an example (see figure 7), if a convolutional layer consists of 16 neurons (e.g., 4X4) and each neuron having the random weight 9 (3X3) +1 bias=10 parameters then this convolutional layer obtained total 160 parameters (16*10). Similarly, the parameters sharing explained the use of the same parameters of the preceding neurons of the convolutional layer to the succeeding neuron of the next convolutional layer. In case of the FC layer, parameter sharing is restricted and used only once (Mboga, 2017). In addition, equivariant representations explained the change of input and output happens in the same way (Goodfellow et al., 2016).

The convolutional layer is followed by non-linear activation (e.g., RELU mentioned in MLP) and pooling (e.g., subsampling). The output volume of the non-linear activation layer is the same as the input volume because the output volume is independent of the non-linear activation. The sub-sampling of the activation layer was done using the max-polling function, which helps to downsampling the input volume and reduces the parameters (Goodfellow et al., 2016). The max-polling function is performed by the function of maximum aggregation algorithm using the *n*-size of pooling region with a specific number of the stride (e.g., 2 X 2 pooling is used as an example in figure 7). Thus, the pooling output layer is the input for the next convolution layer, and again the same function is used for the following pooling layer (e.g., equivariant transformation). The final pooling layer is the input for the fully-connected layer (1 X 1 X depth). At the end of the fully-connected layer, the SoftMax activation was applied to obtain the output class probability as mentioned in MLP. Thus, a wider layer (e.g., convolutional layer) is transformed into a dense layer (e.g., FC) as observed in the CNN architecture (Lee & Kwon, 2016). The CNN network is trained, regularized and optimized through feedforward and backpropagation algorithm with stochastic gradient descent (SGD) as mentioned in MLP.

A supervised CNN network is more relevant to train the image patches from VHR imagery because more complex contextual features can be extracted from the imagery with very high spectral and spatial resolution (Bergado, 2016; Mboga, 2017; Paoletti et al., 2017). In a supervised CNN, the network trains the training patches associated with label information to learn more invariant and complex local contextual features which are validated with the unseen test data (e.g., validation set). As explained the function and architectures of the supervised CNN, the network is incorporated with several hyperparameters such as learning and regularization parameters and hyperparameters related to the network architecture which have explained in the appendix table A1.1. The inconsistent use of hyperparameters in the networks leads to the overfitting of the networks in the learning process. In addition, others parameters like size of training samples and data augmentation are one of the considerations to control the overfitting of the network (CS231n, 2018; Bergado, 2016; Mboga, 2017). The overfitting of the network causes a drop of the overall classification accuracy. In this regard, K-fold cross-validation is one option to mitigate the overfitting of the CNN networks (Bergado, 2016; Mboga, 2017). The hierarchical order of different layers in a supervised CNN (e.g., input, convolutional layers, activation, pooling, and FC) are varied according to the different network architecture (e.g., LeNet, AlexNet, GoogLeNet, VGG Net, etc.) (CS231n, 2018). The simpleCNN wrapper of MatConvNet library provides an efficient and simple CNN architecture with linear chains of a computational building block as compared to other libraries (MatConvNet, 2018).

2.6. Multi-resolution image segmentaion

Image segmentation is a primary step in OBIA for land cover or land use classification (Kohli, 2015; Kuffer et al., 2017; Pratomo, 2016). There are several image segmentation algorithms (chessboard, quadtree, contrast split, etc.). However, multi-resolution segmentation (MRS) is widely used for land cover or land use classification using Ecognition software (Kohli, 2015; Kuffer et al., 2017; Pratomo, 2016). The multi-resolution image segmentation is a bottom-up approach in which the scale parameter (scale size), compactness and shape are an important for aggregating homogeneous object based on their shape, size and compactness in the imagery (Drăguţ & Eisank, 2012; Drăguţ et al., 2014; Pratomo, 2016). The selection of an appropriate scale parameter is a challenging task and depends on the application (Pratomo, 2016). In the most studies, the selection of the scale parameter, shape and compactness is based on trial and error using visual assessment (Pratomo, 2016) (See appendix table A2.1). In recent studies, ESP tool (estimate scale parameter) was efficiently used for estimating scale parameters (Drăgut et al., 2014) while shape and compactness were selected based on the visual fitness of the segmentation (See appendix table A2.1). The ESP tool is chosen for MRS because it estimates the best scale parameters (SP) which control the internal heterogeneity (spectral) of image object that corresponds to their average size (Drăgut et al., 2014).

The ESP tool in MRS is an interactive process which is carried out in a three-level hierarchy (e.g., level1, level2 and level 3) based on the local variance (LV) and rate of change of local variance (ROC/ROC-LV) of multiple layers (e.g. eight spectral bands in this study) at image object level (Drăgut et al., 2014) (see appendix figure A2.1). In this interactive process, the segmentation is hierarchically passing from lower to upper level based on the condition satisfied based on the LV value see appendix figure A2.1. In this process, the ESP tool automatical calculates the local variance, which is explained by equation 19 (Drăgut et al., 2014) while the rate of change of LV is explained by equation 20 (Drăguţ, Tiede, & Levick, 2010).

$$mean\ LV \quad = \quad \frac{LV_1+LV_2+\cdots\ldots\ldots+LV_n}{n} \tag{19}$$

$$ROC \quad = \quad \left[L-\frac{(L-1)}{L-1}\right]\times 100 \tag{20}$$

Where, n=number of image layers, L=LV at the target level, L-1=LV at next lower level. The LV value is increased with the increase of the scale parameter (size of segmented objects) while ROC-LV is opposite to the scale parameter which is explained in appendix figure A2.2. The ROC-LV curve explains how the LV value changes from one object level to another with changing of scale (Drăguţ et al., 2010). According to Drăguţ et al. (2010), only LV cannot help to select the best scale parameters, and in this regard, both LV and ROC-LV are required to select the best scale parameter because both explain the change of LV with the size of the segment. The peak of ROC-LV curve illustrates the best segmented object at an appropriate scale with lesser internal heterogeneity (Drăguţ et al., 2010). The segmentation sometimes suffers from under-segmentation (e.g., exceed the boundary of the target object to others object) or over-segmentation (e.g., multiple segmentation of the same object) (Drăguţ et al., 2010). Drăgut et al.(2014) statistically validated the segmentation based on Area Fit Index (AFI), index of under-segmentation and over-segmentation and Quality Rate (QR). The under-segmentation is more problematic than over-segmentation because multiple segments are easy to merge to get large object while it is hard to split segments (containing several objects).

2.7. Urban land use mapping in India

From literatures on Indian cities, few researchers classified single urban land use from VHR imagery either using OBIA (e.g., Kuffer & Barros, 2011; Sameen & Pradhan, 2016) or combining PBIA and OBIA (e.g., Kuffer et al., 2017) while few authors classified detailed urban land use from coarse resolution imagery using traditional classifier in PBIA (e.g., Wentz et al., 2008). Furthermore, few studies in last decades employed detailed urban land use classification using coarse resolution satellite imagery and visual interpretation (e.g., Pathan et al.,1989; Pathan et al., 1991). In 2006, the Ministry of Urban Development, Government of India initiated the National Urban Information System (NUIS) programme for developing a detailed geospatial database of urban land use at 1:10,000 scale for 152 Indian cities (including Bengaluru city). In this programme, high-resolution satellite imageries are used for visual interpretation in a GIS environment to classify detailed urban land use for urban development and management (NRSA, 2008). The employed method is very slow, expensive and information get quickly outdated.

2.8. Urban land use pattern in Indian cities

The urban land use pattern in Indian cities has a complex urban form, design, and function due to unplanned urban developments (Chadchan & Shankar, 2012; Kotharkar, Bahadure, & Sarda, 2014). The urban form explains the spatial arrangement of different land use across the city while urban design explains the architectural form of the city (e.g., shape, size of buildings, etc.). In addition, the function of land use relates to use of land for a specific purpose such as residential land use for living, commercial land use for trade and commerce, etc. (Alam, 2011). The complex urban form relates to the unorganized form of land use which does not follow the standard urban land use planning model (e.g., sector model, Hoyt, 1939; concentric zone model, Burges, 1925, etc.) (Alam, 2011). In unorganized urban form, shape, size, structure, orientation, colour and function of different urban land uses are often very similar because some of the land use mixed with other land uses (e.g., commercial with residential or commercial with industrial, etc.) (Alam, 2011; Chadchan & Shankar, 2012). This complexity of urban land use in Indian cities was occurred due to the ill implementation of planning regulation by the local planning authority due to political interferences (Chaplin, 2011). Most Indian cities, including Bengaluru, have complex urban land use patterns, which are shown in master planning maps (appendix figure B3.2) and satellite imagery (appendix figure B3.1).

3. RESEARCH METHODOLOGY

In this study, urban land use classification of the Bengaluru city was carried out from the VHR imagery using SVM-RBF and CNN with OBIA and BBIA which explained in the appendix figure B1.1. The land use classification carried out using SVM-RBF and OBIA, or BBIA is termed as SVM-RBF-OBIA or SVM-RBF-BBIA. Similarly, the land use classification carried out using CNN and BBIA is termed as CNN-BBIA. In SVM-RBF-OBIA-based urban land use classification, several sub-processes were employed, i.e., multi-resolution segmentation, features extraction, aggregation and normalization at object-level, feature selection, parameter tuning, classification, validation, and accuracy assessment. Similarly, SVM-RBF-BBIA-based urban land use classification employed block generation (regular grids), features extraction, aggregation and normalization at block-level and others process are the same as for SVM-RBF-OBIA. In addition, in CNN-BBIA several sub-processes were employed such generation of patch-based 3-D training samples, parameter tuning (e.g., learning and regularization parameters, hyperparameters, depth, etc.), classification, validation, and accuracy assessment. Finally, accuracy assessment indices (overall accuracy, kappa, etc.) were compared to assess their performance for all urban land use classification outputs. The above stated methods are explained below:

3.1. Study area

Bengaluru city is located in the Karnataka state, India (see appendix figure B2.1). The Bengaluru city has 8,495,492 inhabitants with an annual population growth rate of 3.25 percent. The city is known as the "IT hub of Asia" and "Silicon Valley of India" covering an area of 786 sq. km built-up land (Bangalore Development Authority, 2007; Census, 2011). The city attracts several national and international investments for developing better trade, commerce, industry and living infrastructures. Timely and accurate information of city growth is required to support competitive economic growth and sustainable urban development (United Nations, 2015). In this regard, Bengaluru city is selected as a study area to map the urban land use from VHR imagery using machine learning algorithms due to non-availability of a detailed land use map. In this study, a small part of the city close to city centre (two tiles covering 850 by 850 metre each tile) is selected because most of the proposed land use classes are observed in this area.

3.2. Datasets and software used

The Worldview3 multispectral (MS) and panchromatic (PAN) VHR satellite imagery and Bengaluru city land use map as the secondary data used in this study (see appendix table B3.1, figure B3.1, and B3.2). The VHR imagery of 15th February 2015 was collected from the DynaSlum project (http://www.dynaslum.com/overview/). The Bengaluru city reference land use map is collected from the revised master plan, 2015. This reference land use maps show details of many land uses such as urban green, vacant land, etc. was missing from the original ground information as observed from the satellite image, Google Earth and Openstreet map (see appendix figure B3.1 and B3.2). A fused VHR imagery (MS+PAN) has been used for extracting image features while the land use map was used for extracting training, validation and test data. In addition, Google Earth and Openstreet map are used for preparing a referenced map because few land use/cover classes like vacant land, urban green, shadow, and waterbody were not in the master plan map. Two tiles, of each 2501 by 2501 pixel were selected to measure the transferability of the selected machine learning algorithms. Several softwares and programming languages were used to carry out this research which explained in the appendix, table B3.2.

3.3. Pre-processing

The pre-processing of satellite imagery and other referenced data was carried out before feature extraction and classification. Firstly, a fusion of MS and PAN was carried out by High Pass Filter (HPF) resolution merge using Erdas Image 2015 (see figure 8). This improves the spectral and spatial resolution of the image. The High Pass Filter image fusion was selected in this study because it provides an excellent detailed and a realistic representation of the scene as compared to others image fusion techniques (e.g., Intensity-Hue-Saturation, Principle Component Analysis, Gram-Schmidt, etc) (Hexagon Geospatial, 2017; Nikolakopoulos & Oikonomidis, 2015; Nikolakopoulos & Konstantinos, 2008; Yusuf et al., 2013).

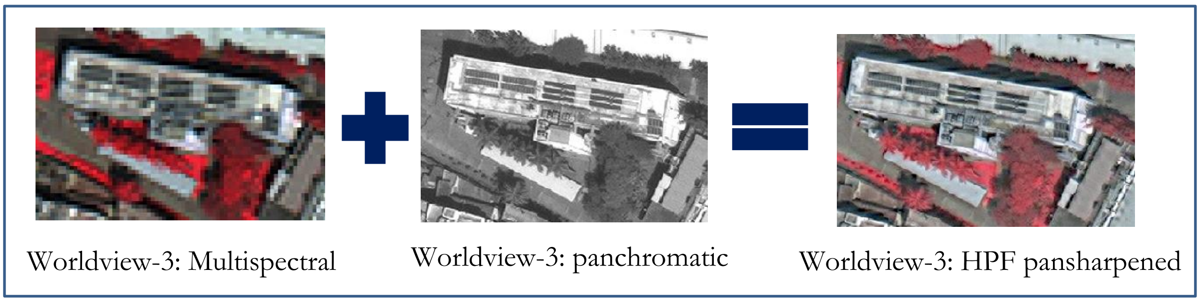


Figure 8. Shows HPF resolutions merge of Worldview-3 multispectral and panchromatic bands, 2015.

Secondly, master plan map was georeferenced with VHR imagery (UTM projection, Zone 43 and datum WGS 1984) using polynomial model one in Erdas Image 2015. Thirdly, based on the proposed land use classification scheme, the referenced land use map was prepared for validation and assessing the accuracy of the classified land use maps. The master plan map of the Bengaluru city, google earth and Openstreet map were used to prepare this reference map using visual interpretation. The minimum mapping unit in the visual interpretation was 100 m^2, which was decided based on the minimum scale size of the selected block (i.e., 29×29 pixel=97.23m^2) in this study. Finally, referenced land cover maps were prepared by aggregating this referenced land use maps for validation and accuracy assessment of the classified urban land cover maps.

3.4. Multi-resolution image segmentation

Based on the literature review the multi-resolution image segmentation (MRS) was used for OBIA. In this study, eight spectral bands of fused VHR worldview imagery were used for segmentation using MRS with EPS tool in eCognition software. The compactness parameter 0.80 and shape parameters 0.50 were used for MRS segmentation followed by bottom-up segmentation approach. MRS-ESP was used to automatically extracted the three optimal scale parameters (e.g., level 1, level 2 and level 3) with user-defined increments of scale size 1 for MRS level 1, 5 for MRS level 2 and 50 for MRS level 3. Validation of the different segmentation level was carried out based on the land use classification accuracies.

3.5. Block generation

In this study, regular square grids are used to aggregate different image features. Three types of block have been selected, block 1 (29×29 pixels=9.86×9.86 metre), block 2 (43×43 pixels=14.62×14.62 metre and block 3 (59×59 pixels=20.06×20.06 metre). The blocks were prepared using fishnet tool in Arc GIS 10.5.1. Block 1 is selected based on the minimum mapping size of land use in the image. It is assumed that the minimum size of the urban green area (e.g., vegetated area), transportation (e.g., medium size road) are closed to 10 metre, and the other land use (e.g., single tree, narrow road, etc.) smaller than 10 metre are excluded. Sandborn & Engstrom (2016) argued that block size 4 (9.76 metre) is more significantly correlated with extracted features than the block size 8 (19.52 metre). Thus, block2 and block3 have been selected with the linear incrementation of approximately 5 metre (14 pixels) for the urban land use classification.

3.6. Feature extraction and aggregation

In this study, image features are conceptualized and selected based on literatures review (see figure 9). The spectral (e.g., mean bands, mean rightness, NDVI), textural (e.g., GLCM, GLDV, LBP, MPPR), geometric (e.g., asymmetry, compactness, Elliptical fit etc.) and contextual features (spatial metrics, e.g., patch density, aggregation index, fractal dimension etc.) were separately extracted and aggregated for OBIA and BBIA (See figure 9). The spectral, textural and geometric features were extracted from fussed VHR imagery while contextual features were extracted from urban land cover (e.g., built-up) using the Fragstats software. In this study, the mean was used for aggregating image features. However, CNN has the self-extraction capability of image features explained in section 3.10.2.

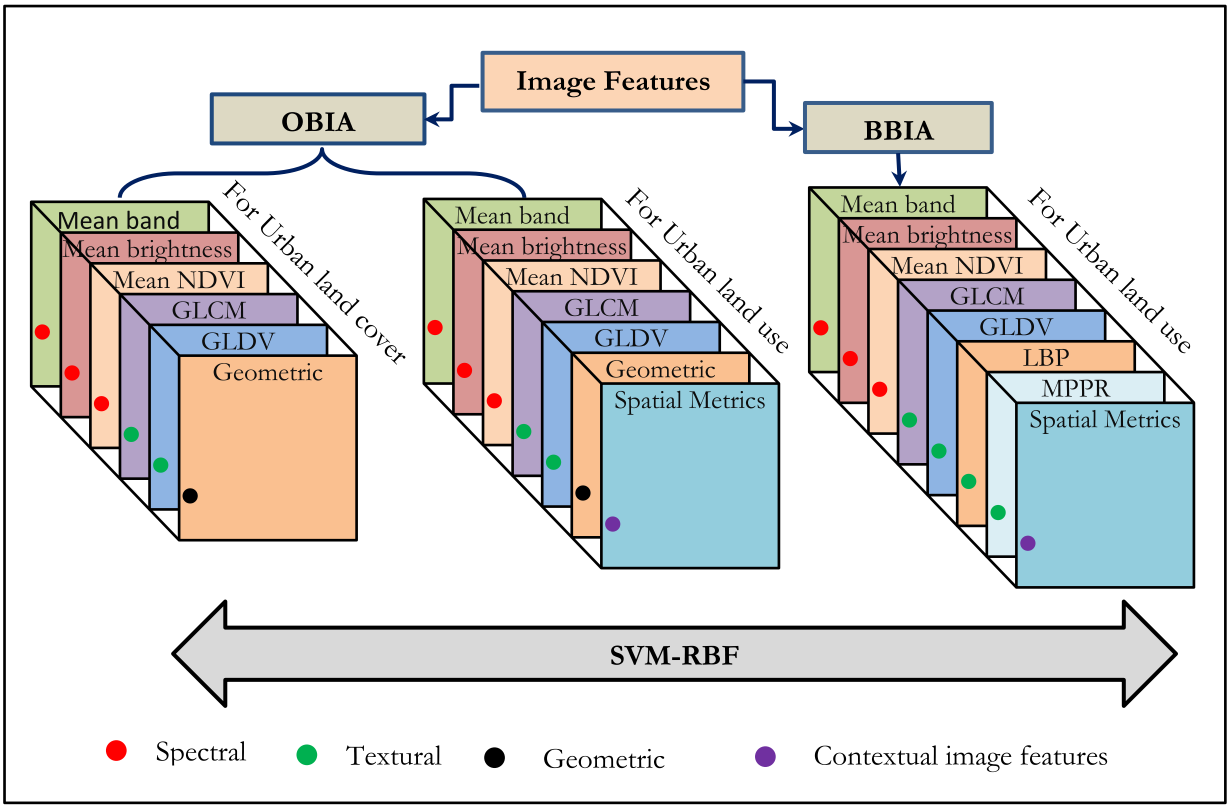


Figure 9. Shows feature selected for OBIA and BBIA for urban land use/cover classification using SVM-RBF.

The image features as summarised in figure 9 for OBIA and BBIA-based urban land cover/use classification using SVM-RBF are explained below:

3.6.1. Features for OBIA

In this study, following image features were extracted and aggregated in OBIA. These object-level features have been used for both urban land cover and urban land use classification:

- **Mean spectral band**

The mean of spectral value (DN values) of all eight spectral bands of the original fussed VHR image was extracted and aggregated, providing the mean spectral response of different urban land cover/use objects (Aguilar et al., 2012).

- **Mean brightness**

The mean brightness was extracted by averaging brightness of 8 spectral bands. The higher mean values indicate brighter objects while lower mean values indicate darker objects (Aguilar et al., 2012). The lower mean brightness value shows the waterbody, shadow, road networks and dark green vegetation in a dark colour while higher brightness values show buildings, light vegetation and vacant land in grey to bright colour. However, the variation of brightness in different land uses such as residential, commercial, industrial is depends on the materials used on the roof, level of pollution and age of buildings, etc.

- **Mean normalized difference vegetation index**

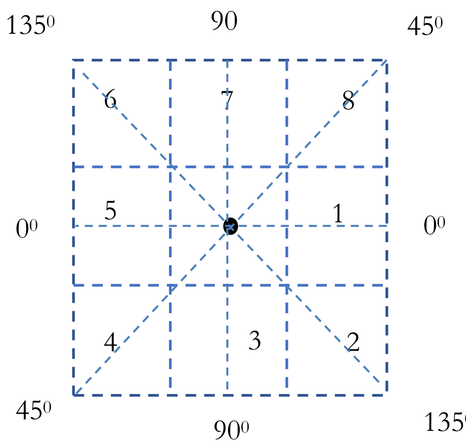
The mean normalized Difference Vegetation Index ($NDVI_{mean}$) is the robust indicator either used for separating different types of vegetation or for separating vegetation from the non-vegetation areas (Nouri, Beecham, Anderson, & Nagler, 2013). The mean NDVI was extracted from the mean Near-Infrared (NIR1, band7) and mean Red (band5) spectral bands using equation 21 (Nouri et al., 2013).

$$NDVI_{mean} = \left(\frac{(Mean\ band_7 - Mean\ band_5)}{(Mean\ band_7 + Mean\ band_5)}\right) \tag{21}$$

A higher mean NDVI value indicates healthy vegetation and lower NDVI unhealthy vegetation or non-vegetation areas (e.g., built-up areas) (Nouri et al., 2013). Thus, mean NDVI is widely used for distinguishing urban from non-urban areas (Nouri et al., 2013) as well as classifying different types of urban land cover/use in VHR imageries (e.g.,Berger et al., 2013; Man et al., 2015; Sandborn & Engstrom, 2016; Silván-Cárdenas et al., 2014; Sun et al., 2016; Zhan, Molenaar, & Xiao, 2001).

- **Gray level co-occurrence vector**

The Gray Level Co-occurrence Matrix (GLCM) is a robust and widely used algorithm for extracting texture from the satellite imagery for characterizing different types land cover/use (e.g., Wieland & Pittore, 2014; Zhang et al., 2017).



Eight gray-tone spatial dependence matrices (GLCM matrices):

- $GLCM_{mean}$
- $GLCM_{standard\ deviation(SD)}$
- $GLCM_{homogeneity}$
- $GLCM_{contrast}$
- $GLCM_{dissimilarity}$
- $GLCM_{entropy}$
- $GLCM_{Angular\ second\ moment(ASM))}$
- $GLCM_{correlation}$

Cell 1 & 5 are 0 degree nearest neighbour, Cell 4 & 8 are 45 degree nearest neighbour
Cell 3 & 7 are 90 degree nearest neighbour and Cell 2 & 6 are 135 degree nearest neighbour

Figure 10. Shows different nearest neighbours and GLCM matrices (Haralick et al., 1973).

The GLCM computes the spatial dependency of grey-tone to characterize the texture of the image objects based on the relationship of angular direction and distances of neighbouring pixel pairs in the image (see figure 10) (Haralick et al., 1973). Haralick et al.(1973) developed 14 GLCM matrices in which eight matrices were extracted using equation 22 to 29 (Laliberte & Rango, 2009) with the user-defined angular direction (e.g., mean of four direction.

$$GLCM_{mean} = \sum_{i,j=0}^{N-1} P_{i,j} \Big/ N^2 \tag{22}$$

$$GLCM_{SD} = \sqrt{\sigma_i^2}; \sqrt{\sigma_j^2} \tag{23}$$

$$where,$$

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-\mu_i)^2 \quad and \quad \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}\,\left(j-\mu_j\right)^2$$

$$GLCM_{homohgenity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{24}$$

$$GLCM_{contrast} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \tag{25}$$

$$GLCM_{dissimilarity} = \sum_{i,j=0}^{N-1} P_{i,j}|i-j| \tag{26}$$

$$GLCM_{entropy} = \sum_{i,j=0}^{N-1} P_{i,j}\left(-In\, P_{i,j}\right) \tag{27}$$

$$GLCM_{ASM} = \sum_{i,j=0}^{N-1} P^2{}_{i,j} \tag{28}$$

$$GLCM_{correlation} = \sum_{i,j=0}^{N-1} P_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \tag{29}$$

Where, P_{ij}, represents normalized gray value at location ij of the matrix; σ_i and σ_j are the standard deviation of the row, i and column, j; μ_i and μ_j are mean of the row, i and column, j and N is the number of row and column. In this study, the above eight GLCM matrices were applied on each of the eight spectral bands of the fussed VHR image. The GLCM matrices were calculated in mean of four angular directions (0^0, 45^0, 90^0, 135^0) because several studies proved that this angular direction provides very optimum results (e.g., Laliberte & Rango, 2009; Shabat & Tapamo, 2014; Yanchen et al., 2014; X. Zhang et al., 2017). These measures help to separate one urban land cover/use from another because each urban land cover/use has distinct spatial identity such as smoothness, orderliness, and orientation, etc.

- **Gray level difference vector**

The Gray Level Difference Vector (GLDV) is the sum of the diagonals of the GLCM, which is used to calculate the absolute difference of neighbours (Aguilar et al., 2012; Laliberte & Rango, 2009; Shabat & Tapamo, 2014). Four GLDV matrices were calculated for each of the eight spectral bands with a mean of four directions (0^0, 45^0, 90^0, 135^0) using equation 30-33 (Laliberte & Rango, 2009).

$$GLDV_{mean} = \sum_{i,j=0}^{N-1} V_k \Big/ N^2 \tag{30}$$

$$GLDV_{contrast} = \sum_{i,j=0}^{N-1} P_k(i-j)^2 \tag{31}$$

$$GLDV_{entropy} = \sum_{i,j=0}^{N-1} P_k \,(-In\, P_k) \tag{32}$$

$$GLDV_{ASM} = \sum_{i,j=0}^{N-1} P^2{}_k \tag{33}$$

Where V_k represents the normalized GLDV and k is equal to |i-j|. Previous researches showed that the classification accuracy improved using the GLDV matrices for urban land cover/use from VHR imagery (e.g., Aguilar et al., 2012; Laliberte & Rango, 2009).

- **Geometric features**

The geometric features (see appendix table B4.1) allow to separate one image object from others based on their shape and size. They are widely used to improve classification accuracy for urban land use classification (Aguilar et al., 2012; Y. Huang et al., 2017; Lei Ma, Cheng, Li, Liu, & Ma, 2015; Yanchen et al., 2014), because each urban land use has a distinct shape, size and spatial arrangements (Sandborn & Engstrom, 2016).

- **Spatial metrics**

The spatial metrics are robust indicators used for quantifying spatial structures and patterns based on density, aggregation, fragmentations, cohesion, and shape of the different spatial objects (Herold et al., 2005; Herold et al., 2003; McGarigal et al., 2012). Spatial metrics were primarily used quantifying dynamics of different land covers (e.g., Herold et al., 2005) but recently, they are increasingly used for urban land use classification (e.g., Herold et al., 2003; Kohli, 2015; Kuffer & Barros, 2011; Kuffer et al., 2017) because, each urban land use is shaped by either aggregation and cohesion or fragmentation of different urban land covers. The built-up land cover (e.g., using OBIA-SVM-RBF classification) map was used for the extraction of the following spatial metrics using equations (see table 1) (McGarigal et al., 2012) in Fragstats 4.2.1. These spatial metrics were selected based on previous studies (e.g., Herold et al., 2003; Kohli, 2015; Kuffer & Barros, 2011; Kuffer et al., 2017), showing an improve in classification accuracy along with spectral, textural and geometric features.

Table 1. Description of the spatial metrics used for OBIA/BBIA-based urban land use classification (McGarigal et al., 2012).

No.	Spatial metrics	Equations	Range	Descriptions
1	Patch Density	$\frac{n_i}{A}(10{,}000)(100)$	PD>0, constrained by cell size	PD measures the number of patch per 100 hectares.
2	Aggregation Index (AI)	$\left[\frac{g_{ii}}{max \to g_{ii}}\right](100)$	$0 \leq$ AI ≤ 100	AI measures the probability of a patch likely to be the same class. AI is equal to 0 when patches are maximally disaggregated and 100 when they are maximally aggregated.
3	Fractal Dimension	$\frac{2\, In\,(0.25\, P_{ij})}{In\, a_{ij}}$	$1 \leqq$ FD $\leqq 2$	FD measures the shape complexity of the patch. FD equal to 1 represents simple shape while 2 represents the highly convoluted shape of the patch
4	Cohesion	$\left[\frac{\sum_{j=1}^{n} P_{ij}}{\sum_{j=1}^{n} P_{ij}\sqrt{a_{ij}}}\right] \cdot \left[1 - \frac{1}{\sqrt{Z}}\right]^{-1} .(100)$	$0 \leq$ Cohesion < 100	Cohesion measures the physical connectedness of the patch. Cohesion is equal to 0 explains the landscape becomes subdivided and disconnected while Cohesion is equal to 100 explains opposite.
5	Largest Patch Index	$\frac{\sum_{j=1}^{a} \max(a_{ij})}{A}(100)$	$0 <$ LPI $\leqq 100$	LPI measures the dominance of the largest patch in comparison of the entire landscape. LPI approaches to 0 explains the dominance of the largest patch in comparison of the entire landscape is increasingly small while LPI equal to 100 explains landscape consists of a single patch.
where, n_i=number patch in class, i in the landscape; A=total area (m^2) of landscape; g_{ii}=number of like adjacencies between pixels of class, i; max-g_{ii}=maximum number of like adjacencies between pixels of class, i; P_{ij}=perimeter (m) of patch ij; a_{ij}=area (m^2) of patch ij; Z=total number of cell in the landscape.				

3.6.2. Features for BBIA

The following features have also been used for BBIA: Mean Spectral bands, Mean Brightness, Mean Normalized Difference Vegetation Index, Gray Level Co-occurrence Matrix, Gray Level Difference Vector and Spatial Metrics. These features (except spatial metrics) were extracted and aggregated in block-level using chessboard segmentation in E-cognition 9.2 (Kamal, Phinn, & Johansen, 2015; Pedersen, 2016). In this process, weight was given to fishnet grid while scale size is considered same as the size of the fishnet grid. In this study, automatically generated chessboard grid was not used due to resolve the locational mismatch between fishnet grid and chessboard segmentation grid. In addition, spatial metrics were extracted and aggregated in block-level using a user-defined grid (e.g., the regular grid is considered as the user-defined grid) in Fragstats 4.2.1. Beyond these features, some more features have been used for BBIA which as follows:

- **Rotation invariant local binary pattern**

The Local Binary Pattern (LBP) is a very important texture mapping algorithm which was initially used for computer vision, pattern recognition and signal processing (e.g., Pietikäinen et al., 2011; Wang et al., 2014) but recently it is increasingly used for satellite image processing for urban land cover/use mapping (e.g., Sandborn & Engstrom, 2016). Primarily, regular LBP ($LBP_{P,R}$) was used to extract the texture from the image based on the monotonic transformation (e.g., linear) of gray value in a circular symmetric neighbour set of pixels in the neighbourhood. This was calculated by thresholding and weighting using equation 34-38 (Wang et al., 2014).

$$LBP_{P,R}\ (x_c, y_c) = \sum_{P=0}^{P-1} s(\ g_P - g_c)2^P \qquad (34)$$

$$and, s(x) = \begin{cases} 1 & x \quad \geq 0 \\ 0 & x \quad < 0 \end{cases}$$

Where, s(x)=thresholding scale function, x=(g_p-g_c), x_c, y_c=x, y position of the central pixel, g_p=pixel value of the neighbourhood, g_c=central pixel value of the neighbourhood, P=total number of neighbours, R=radius of the neighbourhood. In addition, an optimum size or radius of the neighbourhood was estimated by using the equation 35 and 36 to design an appropriate LBP for texture mapping.

$$Neighbourhood\ size\ (n) = \big((R \times 2) + 1\big) \qquad (35)$$

$$Neighbourhood\ radius\ (R) = \big((n-1)/2\big) \qquad (36)$$

Thus, the diagonal distance of neighbours from the central pixel was calculated using bilinear interpolation method which explained in the equation 37 (Wang et al., 2014).

$$Location\ (g_P) = \left(x_c + Rsin\left(\frac{2\pi i}{P}\right)\right), \left(y_c + Rsin\left(\frac{2\pi i}{P}\right)\right) \qquad (37)$$

Similarly, the regular $LBP_{P,R}$ patterns or bins was estimated for different P values (e.g., 256 for P=8) using the equation 38 (Wang et al., 2014).

$$LBP_{P,R}\ \ pattern/bins = 2^P \qquad (38)$$

Secondly, regular $LBP_{P,R}$ was used to measure the rotated invariant LBP ($LBP^{ri}_{P,R}$) to resolve the issues of rotation. In this regard, a unique identifier was assigned to each rotation invariant local binary pattern using equation 39 (Wang et al., 2014).

$$LBP^{ri}_{P,R} = min\big\{ROR\big(LBP_{P,R}, P\big) \mid \ \ P = 0,1, \dots \dots, P-1\big\} \qquad (39)$$

Where, r_i=rotated invariance, ROR (x, P) explained the $LBP_{P,R}$ code, x is rotated P times in clockwise direction at 45^0 angular intervals around the central pixel. In this study, $LBP^{ri}_{8,14}$, $LBP^{ri}_{8,21}$, and $LBP^{ri}_{8,29}$ were used to extract the image texture for eight spectral bands based on the different radius (R=14, 21, 29). This radius was selected based on the chosen block size (e.g., 29, 43, 59) in this study. The $LBP^{ri}_{8,14}$, $LBP^{ri}_{8,21}$, and $LBP^{ri}_{8,29}$ were extracted a combination of 36 unique rotated invariant uniform (e.g., two patterns as 0-1 and 1-0) and non-uniform (e.g., combination 0-1 pattern) spatial transition patterns using the Matlab code developed by Nikolay S. (2017). In this study, P=8 and different R for $LBP^{ri}_{P,R}$ were selected because, in many studies (e.g., Mehta & Egiazarian, 2013; Ojala, Pietikainen, & Maenpaa, 2002), P=8 and different R (R>=1) produce higher texture classification accuracy as compared to the other combination of P and R. Finally, this feature was aggregated in fishnet grid using zonal attributes of Erdas Imaging software.

- **Morphological profiles for partial reconstruction**

The mathematical morphological profiles (MPs) were initially developed for the pattern recognition, face recognition and computer vision (e.g, Dias, Cousty, & Najman, 2014; González-Castro, Debayle, & Pinoli, 2014) but recently, this is widely used for mapping urban land cover/use from VHR imagery (e.g., Dalla Mura et al., 2010; Liao et al., 2015, Liao et al., 2017). MPs were used to extract shape geometry of image object based on the concatenation of morphological closing ($\prod_{\varphi}$) and morphological opening profiles ($\prod_{\gamma}$) using structural element and partial reconstruction operator (e.g., partial geodesic) explained in equation 40 (Dalla Mura et al., 2010).

$$MP_s(f) = \prod_i : \begin{cases} \prod_i = \prod_{\varphi_\lambda}, & with \quad \lambda = (n-1+i), \quad \forall\lambda \in [1, n]; \\ \prod_i = \prod_{\gamma_\lambda}, & with \quad \lambda = (i-n-1), \quad \forall\lambda \in [n+1,\ 2n+1]. \end{cases} \tag{40}$$

The morphological opening profile (γ_R) was used to removing brighter connected object from the image (*f*) with erosion (ε^i) followed by dilation (δ^i) operator. In this regard, same size of the structural element (i) and partial geodesic reconstruction operator by dilation (R^{δ}_f) was used which explained in equation 41 (Dalla Mura et al., 2010).

$$\gamma_R^i(f) = R_f^{\delta}\left(\varepsilon^i(f)\right) \tag{41}$$

Similarly, the morphological closing profile (φ_R) was also used to remove the darker connected object from the image (*f*) with dilation (δ^i) followed by erosion (ε^i) operator. In this regard, same size of the structural element (i) and partial geodesic reconstruction operator by erosion (R^{ε}_f) was also used which explained in equation 42 (Dalla Mura et al., 2010).

$$\varphi_R^i(f) = R_f^{\varepsilon}\left(\delta^i(f)\right) \tag{42}$$

In this study, partial geodesic reconstruction was selected because it can preserve the actual shape and size of the rectangular or near rectangular object as compared to geodesic reconstruction while employing disk shape SE (Liao et al., 2017). The partial geodesic reconstruction was computed measuring partial geodesic distance using the equation 43.

$$Partial\ geodesic\ distance\ (d) = 2(\sqrt{2-1})R \tag{43}$$

However, in MPPR one limiting factor is the shape (e.g., square, disk, etc.) and size of the SE because an object smaller than the SE is automatically deleted and with an increase in size of SE more objects are deleted (Dalla Mura et al., 2010; Liao et al., 2017). In this study, disk shape SE was used because it is very commonly used for urban land use mapping from the VHR imagery (e.g., Liao et al., 2017). The radius (R) and size of SE were computed using the equation 44 and 45.

$$Radius\ of\ SE\ (R) = \frac{S+1}{2} \tag{44}$$

$$Size\ of\ SE\ (S) = (2R\ \ -1) \tag{45}$$

In this study, morphological profiles with partial reconstruction (MPPRs) with disk shape SE was used to extract morphological features from the panchromatic image using the Matlab code developed by Liao et al., (2017). The SE size was considered same as the selected block size 29, 43 and 59 as defined in this study. Finally, this feature was aggregated in fishnet grid using zonal attributes of Erdas Imaging software.

3.7. Normalization of image features

The extracted image features and fused image were normalized using max-min standardization method using ENVI classic 5.3. The value of the normalized image features varies from 0 to 1. The normalized image features were used for feature selection, parameter tuning and land cover/use classification using SVM_RBF. The normalized fused image was used to train the CNN for land use classification.

3.8. Urban land cover/use classification scheme

In this study, six relevant urban land covers and seven urban land uses are selected based on literatures review, national (NRSA, 2008), and local land cover/use classification schemes (e.g., master plan) and empirical experience of the study area (see figure 11). In the local land use classification scheme 9 main land use and 15 sub-land use classes were mapped, but in this study area, only 6 main land use and 12 sub-land use classes are observed. In this study, few land use classes were merged into main land use class because of their homogeneous urban characteristic (appendix table B5.1). In addition, few land use classes have been added in this study due to missing of such land use classes in the master plan map (see appendix figure B3.2). In this study, shadow is not a land cover/use but has included to understands the ability of the classifiers (or classification approach) to mitigate or extract the shadow in final classification because VHR image has a big issue of shadow effect. Therefore, the proposed urban land cover/use has explained in the appendix table B5.1.

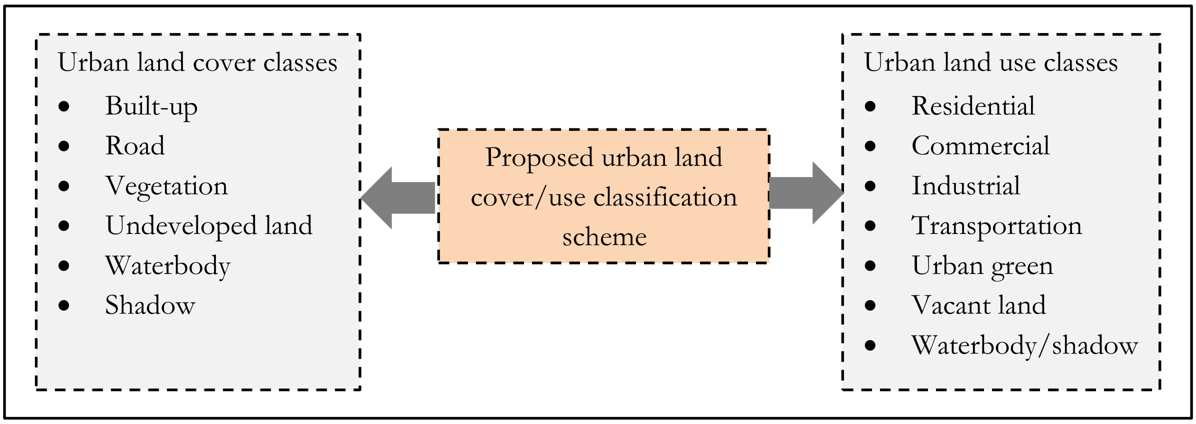


Figure 11. Shows proposed urban land cover/use classification scheme.

3.9. Selection of training and test samples

The selection of optimum sample sizes is a very challenging task in image classification to reduce the effects of Hughes phenomena (Persello & Bruzzone, 2010; Persello & Bruzzone, 2016). In this study, segment/block was selected as the training sample unit while pixel as the test sample unit which was labeled using referenced land cover/use maps. In previous studies several sampling techniques (see equation 46 and 47) were employed to select the optimum size of training samples (Park & Stenstrom, 2008).

$$N = \frac{4p(100 - p)}{\varepsilon^2} \qquad (46)$$

$$N = 30nc \qquad (47)$$

Where, p=expected accuracy (%), ε=allowable error (%), n=number image features, c=number of LULC classes. However, in this study user-defined training samples was selected using stratified random sampling (seeding 1002) depending on the available pure sample segments or blocks (e.g., block with single land cover/use) (see figure 12). The number of pure segments/blocks are controlled by the scale size (e.g., Yanchen et al., 2014; Zhang et al., 2017; Zhen et al., 2013). In addition, training samples were randomly split into training and validation sets using 60:40 rule (e.g., Duque, Patino, & Betancourt, 2017) for features selection and parameter tuning (Persello & Bruzzone, 2010). However, in CNN training samples were increased (e.g., augmented training samples) by rotating original training samples in different angles. The test set was selected as whole tiles (6,255,001 pixels) to assess both pure and mixed segments or blocks for pixel-based final accuracy assessment (see figure 12). The strategy for selecting training, validation and test samples are explained in figure 12.

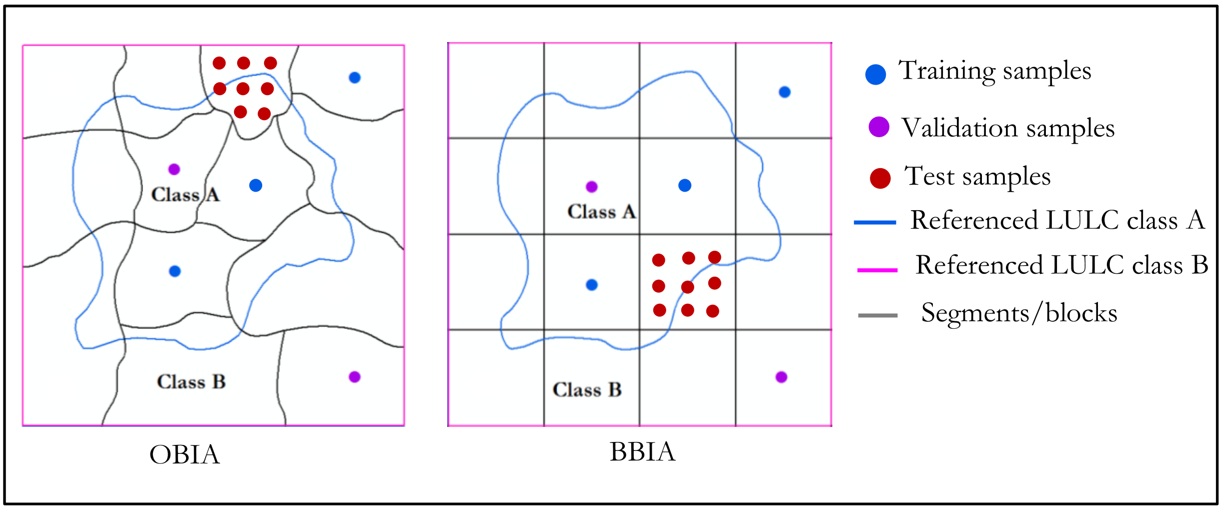


Figure 12. Strategy for selection of training, validation and test samples for OBIA and BBIA using SVM and CNN.

3.10. Design and implementation of selected machine lerning algorithms

Based on the literature review, SVM-RBF and CNN were implemented in this study for OBIA and BBIA-based urban land use classification which explained below:

3.10.1. Support vector machine with radial basis function

The Support Vector Machine with Radial Basis Function (SVM-RBF) was applied to solve multi-class classification problems. In addition, the cost and gamma parameters of SVM-RBF were tuned using k-fold cross-validation to mitigate the change of overfitting of the classifier and improve the classification accuracy (Persello & Bruzzone, 2014). The parameter tuning is explained in the section 3.12.1. Therefore, the SVM-RBF was developed in R studio programming language for classifying OBIA and BBIA-based urban land cover/use classification.

3.10.2. Convolutional neural networks

As explained in the section 2.5.3, the CNN is configured with Input Layer-Convolutional Layer-Activation Layer-Pooling Layer-FC layer-Output layer which was implemented in this study employing simpleCNN wrapper of MatConvNet in Matlab programing language. As the CNN holds the function (input, activation, feedforward, back propagation, etc.) and structure of MLP (hidden layers) then it is initialized, regularized and optimized with feedforward and backpropagation algorithm with stochastic gradient descent (SGD) (see section 2.5.1 and 2.5.3). Table 2 shows the commonly used initialization and regularization parameters to train the network for urban land use classification which were initially implemented in this study based on the literature review (e.g., Bergado, 2016; Mboga, 2017).

Table 2. Initial learning and regularization parameters.

Hyper parameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.01
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0, 0.5)

In addition, table 3 shows the initial architecture of the CNN which were implemented in this study based on the literature review (e.g., Bergado, 2016; Mboga, 2017). The architecture of the CNN is explained below:

- **Input layer**

In this study, 3D input layers consist of 1000 2D training patches of 29X29 size from eight spectral bands were initially implemented to train the network using SGD (see table 3). This training patches (e.g., training samples) were randomly split into training and validation set using 60:40 rule to train and validate the network.

Table 3. Initial CNN configuration and hyperparameters.

Hyper parameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in $C_{1\text{-}2}$ and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

- **Convolutional layers**

In this study, two convolutional layers with eight neighbourhood filters of 5X5 size (e.g., receptive field) (e.g., eight filters in each of the convolutional layers) were initially implemented to learn the features from input layer using SGD (see table 3). In addition, commonly used stride one was implemented in this layer (CS231n, 2018; Bergado, 2016; Mboga, 2017). Thus, zero padding (e.g., initially 2) was selected based on the equation 18 in section 2.5.3. However, the spatial size of the output volume in convolution layers was managed by the equation 17 in section 2.5.3.

- **Activation and pooling layers**

In this study, a non-linear activation function such as RELU was implemented in the output volume of the convolutional layers to generalize the network using SGD (see table 3). The RELU activation function is explained in equation 12 in section 2.5.2. In addition, the max-pooling function was employed in the output volume of the activation layer for sub-sampling and parameter reduction which helps to extract more abstract features (Goodfellow et al., 2016). In this study, commonly used max-pooling region 2X2, stride two and zero padding one were employed in activation layer to train the network using SGD (CS231n, 2018; Bergado, 2016; Mboga, 2017) (see table 3).

- **Fully-connected and output layers**

Initially, FC layer one with 128 neurons were implemented in this study (see table 3). In addition, seven output layers were implemented to classify the seven urban land use classes. In FC layer, stride one and no zero padding were commonly used (e.g., Bergado, 2016; Mboga, 2017). The posterior probability of this output layer was computed using the SoftMax activation function using equation 13 in section 2.5.2. Thus, cross-entropy objective loss function was implemented in this network to estimate the misclassification error, which was minimized using backpropagation with stochastic gradient descent (SGD) optimization function. The cross-entropy objective loss function and SGD employ equation 14 and 15 and in section 2.5.2. Finally, the posterior probability of output layers was used for land use classification which was used for final accuracy assessment using whole tiles (6,255,001 pixels). However, different learning and regularization parameters and CNN hyperaerated were used for parameters tuning using K-fold cross-validation which explained in section 3.12.2. This helps to select best hyperparameters to reduce the overfitting of the network and improve the overall classification accuracy (Bergado, 2016; Mboga, 2017).

3.11. Feature selection

Feature selection was carried out for SVM-RBF-OBIA and SVM-RBF-BBIA-based urban land cover/use classification because to reduce the risk of Hughes phenomena (Damodaran et al., 2017; Persello & Bruzzone, 2016). SFS-HSIC, a supervised feature selection method was employed because it is very simple, fast, robust as compared to the other feature selection methods (e.g., PCA, rank, etc.) for selecting features from the high-dimensional feature space (Persello & Bruzzone, 2016; Damodaran et al., 2017; Huang et al., 2017). Firstly, Hilbert–Schmidt Independence Criterion (HSIC) was used to estimate the class separability by summarizing the reproducing kernel Hilbert space (RKHS) for class dependency (or similarity) measure (Persello & Bruzzone, 2016; Damodaran et al., 2017). HSIC is the square of the Hilbert–Schmidt norm of the cross-covariance operator ($\|C_{XY}\|^2_{HS}$) which was measured from the RKHS using equation 48 (Persello & Bruzzone, 2016).

$$HSIC\big(H,G,P(X,Y)\big) = \|C_{XY}\|^2{}_{HS} \tag{48}$$
$$where, \|C_{XY}\|^2{}_{HS} = E_{xx'yy'}[k(x,x')l(y,y')] + E_{xx'}[k(x,x')]E_{yy'}[k(y,y')] \\ -2E_{xy}[E_{x'}k(x,x')]E_{y'}[l(y,y')]]$$

Where, $E_{xx', yy'}$ is the expectation over both (x,y) according to the joint probability distribution, P(X,Y) and an additional pair of variables (x',y') with the distribution P(X',Y') drawn independently from the RKSH. Similarly, k(x,x') is the kernel function (Gaussian radial basis function) which is used to evaluate the similarity between input instances while l(y,y') is the kernel function for output instances. The Gaussian radial basis function is explained equation 5 in section 2.4. With the given training set (X,Y), the empirical measures of the HSIC were used to evaluate the class dependency based on the degree of alignment of the input kernel matrix, K and output kernel matrix, L using equation 49 (Persello & Bruzzone, 2016).

$$\widehat{HSIC}(X,Y) = \frac{1}{m^2}TR(KHLH) \tag{49}$$

where, m=number of training samples, H=centering matrix, T=trace operator. The HSIC value is equal to 0 explained the X (e.g., image features) and Y (e.g., land cover/use class labels) variables are highly independent while a higher value of HSIC described the strong dependence between X and Y variables (see appendix figure B6.1). As the RBF kernel used for computing HSIC, so the variation of sigma value affects the HSIC value. Hence, 10-fold cross-validation of sigma value (sigma= 1^{-1} to 10^2) was used to select best HSIC value corresponding to the maximum sigma value. Secondly, Sequential Feature Selection

(SFS) strategy was employed to select the best set of features sequentially which corresponds to the maximum HSIC and Sigma value. The SFS-HSIC was developed in R studio programming language. In this study, 200 training samples were split into training and validation sets using 60:40 rule for selecting and validation best images features. Thus, 120 image features were used for selecting best features set for OBIA-based urban land cover classification using SFS-HSIC. In addition, 125 images were used selecting best features set for OBIA-based urban land use classification while 121 image features were used for selecting best features set for BBIA-based urban land use classification. Finally, SVM-RBF with a fixed cost equal to 100 and best gamma (e.g., Sigma) was used for final accuracy assessment of the selected best features set using 6,255,001 test pixels.

3.12. Parameter tuning

The parameter tuning is a very crucial step in image classification to developed best parameter of classifiers which provided maximum classification accuracy. The parameters tuning was employed for tile1, and same parameters were used for urban land cover/use classification for tile1 and tile2 to assess the domain adaptation (transferability) of the classifiers. This strategy was applied for the SVM-RBF-based OBIA and BBIA and CNN-based BBIA which has explained below.

3.12.1. Parameter tuning for SVM-RBF

In this study, the grid search hold-out 10-fold cross-validation was employed to obtain best cost and gamma parameters of SVM-RBF for OBIA and BBIA-based urban land cover/use classification. The grid search hold-out k-fold cross-validation was applied in this study, because this is a fast, robust and widely used algorithm for parameters tuning in SVM-RBF-based image classification (Damodaran et al., 2017; Li et al., 2016; Liao et al., 2017; Sun et al., 2016). In the study, the cost range (C) is 10^{-1} to 10^{3} (0.1-1000) while gamma (σ) range is 10^{-3} to 10^{1} (0.001 to 10) were selected because the higher value of cost and gamma parameters leads to overfitting in the classification problem (Duque et al., 2017). The cost and gamma range were divided into 10 sequential folds for grid search hold-out cross-validation using SVM-RBF. In this study, parameter tuning was carried out by splitting 200 training samples into training and validation sets to train and validate the algorithm for selecting best parameters. The final overall accuracy of the selected best parameters was assessed using 6,255,001 test pixels (e.g., whole tile).

3.12.2. Parameter tuning for CNN

The hyperparameters of CNN as divided into two categories such as learning and regularization parameters and hyperparameters related to configuration which were used for parameter tuning. As mentioned above, parameter tuning using k-fold cross-validation is a robust approach to resolve the risk of overfitting of the classification algorithm. In this study, 10-fold cross-validation (K) of the hyperparameters was carried which mentioned in table 4, 5 and 6 following the guideline of the previous researches (e.g., Bergado, 2016; Mboga, 2017). In this study, 1000 training samples were split into training and validation sets using 60:40 rule to train and validate the network for selecting best hyperparameters. The final overall accuracy of the selected best hyperparameters was assessed using 6,255,001 test pixels (e.g., whole tile).

- **Learning and regularization parameters**

Table 4 shows the different learning and regularization parameters which were used for 6-fold cross-validation. In this cross-validation process, initial CNN configuration was used which explained in table 5.

Table 4. Experiments on leaning and regularization parameters.

Hyperparameters	Values
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.01, 0.001
Weight decay, λ	0.01, 0.001, 0.0001
Dropout rate, d_r in (D1 & D2)	(0, 0.5), (0.25, 0.5) (0.5, 0.5)

Initially, two convolutional layers and one fully-connected layer were used to train the CNN with stochastic gradient descent (see table 5). A batch size 10 and maximum epoch 1000 were selected based on the initial experiment on others batch size 5, 15 and epoch 500 and 1500 based on classification accuracy on the validation set. It was observed that the network gets overfitted while using batch size 5, 15 and epoch 500 and 1500 as compared to batch size 10 and epoch 1000. In addition, momentum is 0.90 is commonly used to train the network (e.g., Bergado, 2016; Mboga, 2017). However, different learning rate, weight decay, and dropout rate experimented because the network is affected by the risk of overfitting due to inappropriate use of such learning and regularization parameters (Bergado, 2016; Mboga, 2017; Srivastava et al., 2014).

- **Hyperparameters related to CNN configuration**

Table 6 shows different hyperparameters related to the CNN configuration which were tuned using the best learning and regularization parameters and initial CNN configuration showing in table 4 and 5. This initial CNN configuration as showing in Table 5 was selected based on the guideline of the previous research carried by Bergado, 2016 and Mboga, 2017. A patch size 29 was selected based on the highest accuracy obtained by the SVM-RBF-BBIA for urban land use classification. In this study, 4-fold cross-validation (K) was carried out to select the best hyperparameters related to the CNN configuration using the same training, validation and test samples as used for tuning the learning and regularization parameters.

Table 5. Initial CNN configuration.

Hyperparameters	Values
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in C_{1-2} and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolution layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

The cross-validation of these hyperparameters related to CNN configuration is essential because the network is overfitted due to high volume of parameters sharing and complex depth of the network which controlled by the hyperparameters mentioned in table 6 (Bergado, 2016; Mboga, 2017).

Table 6. Experiments on hyperparameters related to CNN configuration.

Hyper-parameters	K-1	K-2	K-3	K-4
Patch size	29	29	29	29
Number of filters	8,16,32,64	Best	Best	Best
Filter dimension	5	5,15,25	Best	Best
Number of convolutional layer (Cn)	2	2	1, 2,3,4	Best
Number of fully-connected layers (FCn)	1	1	1	1,2,3

The varying zero padding (e.g., 2 for filter size 5) in the convolutional layer was used which determined with the equation 18 in section 2.5.3 while the stride one was commonly used. However, the spatial size of the output volume in convolution layers was managed by the equation 17 in section 2.5.3. In addition, pooling size two, zero padding one and stride two were commonly used in the max-pooling layer. Similarly, stride one and no zero padding were commonly used in FC layer (e.g., Bergado, 2016; Mboga, 2017).

3.13. Urban land cover/use classification

The SVM-RBF and CNN were used to perform supervised OBIA and BBIA-based urban land cover/use classification, which is explained below.

- **Urban land cover classification**

The SVM-RBF-based supervised OBIA was used for urban land cover classification using the best selected image features, best cost and gamma parameters and best size of training samples. The classification of the urban land cover is carried out to extract contextual features using spatial metrics.

- **Urban land use classification**

SVM-RBF and CNN were used for supervised OBIA and BBIA-based urban land use classification. The SVM-RBF was used for supervised OBIA and BBIA-based urban land use classification using best selected features, best cost and gamma parameters and best size of training samples. Similarly, the CNN was to carry out supervised BBIA-based urban land use classification using best hyperparameters and best size of training samples.

3.14. Validation and accuracy assessments

The validation allows to test the ability of the classifiers/algorithms to solve the classification problems with reference to ground reality. In image analysis, accuracy assessment of the classification results can be done in many ways (e.g., pixel-based, point-based, area-based etc.) depending on the types of image classification approaches (e.g., PBIA, OBIA, etc.). In PBIA, pixel-based accuracy assessment is most commonly used (e.g., Persello & Bruzzone, 2010) while area-based accuracy assessment is common for OBIA (e.g., Ma et al., 2015) and BBIA (e.g., Duque et al., 2017). The point-based and area-based accuracy assessment approaches using samples test point and polygon (using sampling technique) are more susceptible to the risk of biasness for OBIA and BBIA. This is happened because there is a combination of pure and mixed objects or blocks in OBIA and BBIA occurred due to scale issue (e.g., Ma et al., 2015) which cannot equally be assessed using sampling technique. In this regard, pixel-based accuracy assessment approach was used to evaluate whole tiles pixel-wise (e.g., 6,255,001 test pixels) for OBIA and BBIA-based urban land cover/use classification. The pixel-based accuracy assessment was performed to measure the different quantitative accuracy assessment indices (e.g., overall accuracy, user accuracy, producer accuracy and kappa) from the confusion matrix using equation 50-53 (Persello & Bruzzone, 2010; Huang et al., 2017).

$$Overall\ Accuracy, OA = \left(\frac{\sum_{i=1}^{n} C_{ii}}{N}\right) \times 100 \tag{50}$$

$$User\ Accuracy, UA = \frac{C_{ii}}{C_{i+}} \times 100 \tag{51}$$

$$Producer\ Accuracy, PA = \frac{C_{ii}}{C_{+i}} \times 100 \tag{52}$$

$$Kappa = \frac{N \sum_{i=1}^{n} C_{ij} - \sum_{i=1}^{n} C_{i+}.C_{+i}}{N^2 - \sum_{i=1}^{n} C_{i+}.C_{+i}} \tag{53}$$

Where, n=number of land cover or use classes; N=total number of test samples, C_{ii}=number of correctly classified by the class, i; C_{i+}=row total of class, i; C_{+i}=column total of class, i. The overall accuracy explained the land cover/use correctly classified by the classifier with reference to the test samples and classification output itself. In addition, user accuracy explained the accuracy with reference to classification output itself while produced accuracy explained the accuracy with reference to the test samples. In addition, kappa coefficient +1 means the classification is better than the random classification while 0 and -1 explained the opposite (Humboldt State University, 2018).

3.15. Evaluation of performance

The performances of the selected machine learning algorithms for urban land use classification were evaluated based on the different accuracies assessment indices (e.g., Bergado, 2016). These accuracies assessment indices were measured from the confusion matrix which explained below.

- **Performance assessment based on quantitative indices**

The quantitative performance assessment was done based on the commonly used quantitative accuracy assessment indices such as overall accuracy, kappa, recall, precision, and F1-score. However, recall explained to user accuracy (UA) while precision explained to producer accuracy (PA) (Radoux & Bogaert, 2017). The F1-score explained the harmonic mean of the precision and recall. The overall accuracy, kappa, user accuracy and producer accuracy already explained in the section 3.14. In addition, the recall, precision, and F1-score are explained in percent using the equation 54-56 (e.g., Bergado, 2016).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negetive} \times 100 \tag{54}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100 \tag{55}$$

$$F1 - score = \left(2 \times \frac{Precision \times Recall}{Precision + Recall}\right) \times 100 \tag{56}$$

- **Performance assessment based on locational indices**

The locational performance assessment was done based on the commonly used locational accuracy assessment indices such as Klocation and Kno (Ahmed, Ahmed, & Zhu, 2013; Megahed et al., 2015). The Klocation (kappa for location) explained the kappa agreement for location while Kno (kappa for no information) explained the overall agreement in terms of quantity and location (Eastman, 2012). These indices were measured using validation module in Idrisi Selva. The locational accuracy assessment indices explained classification agreement based on the location at grid cell level.

- **Performance assessment based on other indices**

Beyond, the quantitative and locational agreement, other indices such as quality of the classified map and time taken to execute the classification by the classifiers. The quality of the classified maps was compared with the referenced and local land use classification scheme maps based on the visual interpretation of shape, size, orientation, misclassification of the classified land use. In addition, the time taken for feature extraction, feature selection, and parameter tuning was also compared for assessing overall time taken by the classifiers to complete the classification.

3.16. Domain adaptation

The domain adaptation is a very efficient measure to assess the domain adaptability of the classifiers based on the parameters extracted from one tile and same parameters are used for different tile (Bergado, 2016). In this study, the sampled domain adaptation was employed to classify the urban land use of the tile one from where training samples were taken. In addition, unsampled domain adaptation was also employed to classify the urban land use of the tile two from where training samples were not taken, but the training samples were taken from another tile one.

4. RESULTS AND DISCUSSIONS

This chapter provides the results and discussion. The results are obtained from the selected methods and datasets carried out in chapter 3, to address the research objectives and research questions.

4.1. Generated fused satelliete imagery

Figure 13 shows the fused image of the MS and PAN bands of Worldview-3 imagery which have been produced using the HPF resolution merge (section 3.3). The figure 13 shows that the HPF provides very realistic result without losing the spatial and spectral information of the original satellite image. This fused satellite imagery provides detailed information of the different urban land cover/use as compare to the original satellite imagery (see appendix figure B3.1).

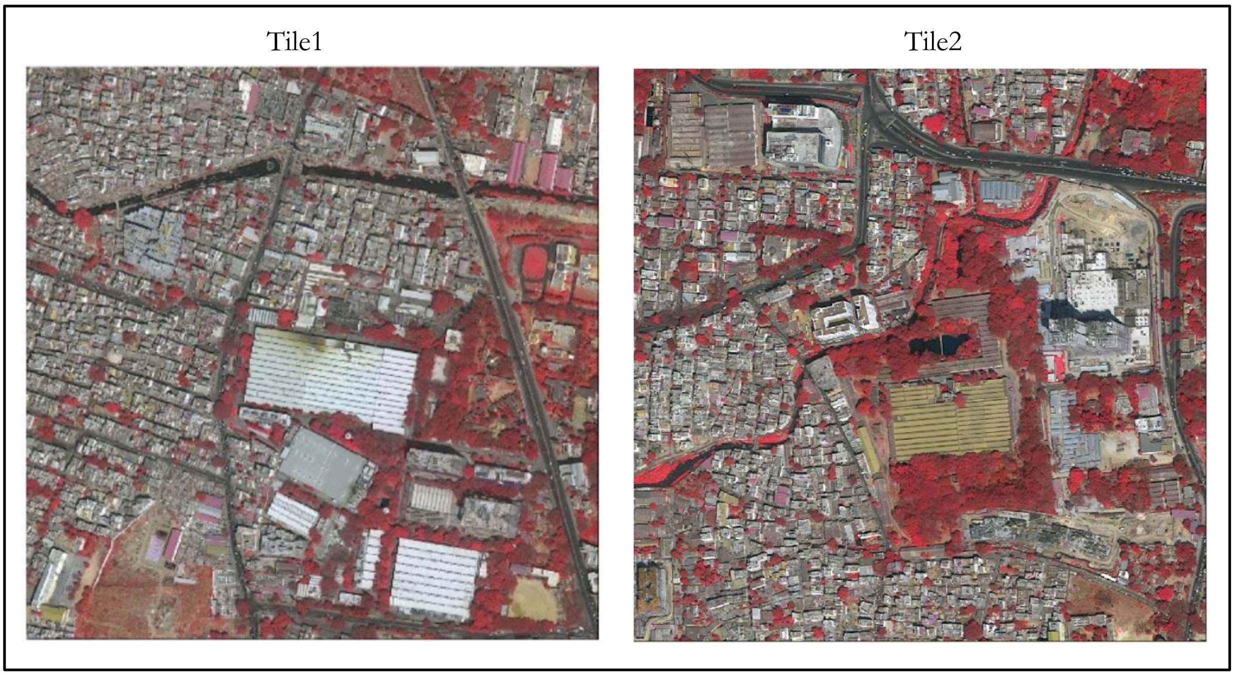


Figure 13. Shows Worldview3 fused satellite imageries of tile1 and tile2 with 2501 × 2501 pixels, 2015.

4.2. Prepared referenced urban land cover/use maps

The figure 14 shows the reference urban land use maps have been prepared as explained in section 3.3. This reference urban land use maps have been used for validation and accuracy assessment of the classified urban land use maps. In addition, reference urban land cover maps that prepare from reference land use map have been used for validation and accuracy assessment of the classified urban land covers maps (appendix figure C1.1). Therefore, each of the referenced land use/cover maps has 2501 row by 2501 column with 6,255,001 pixels.

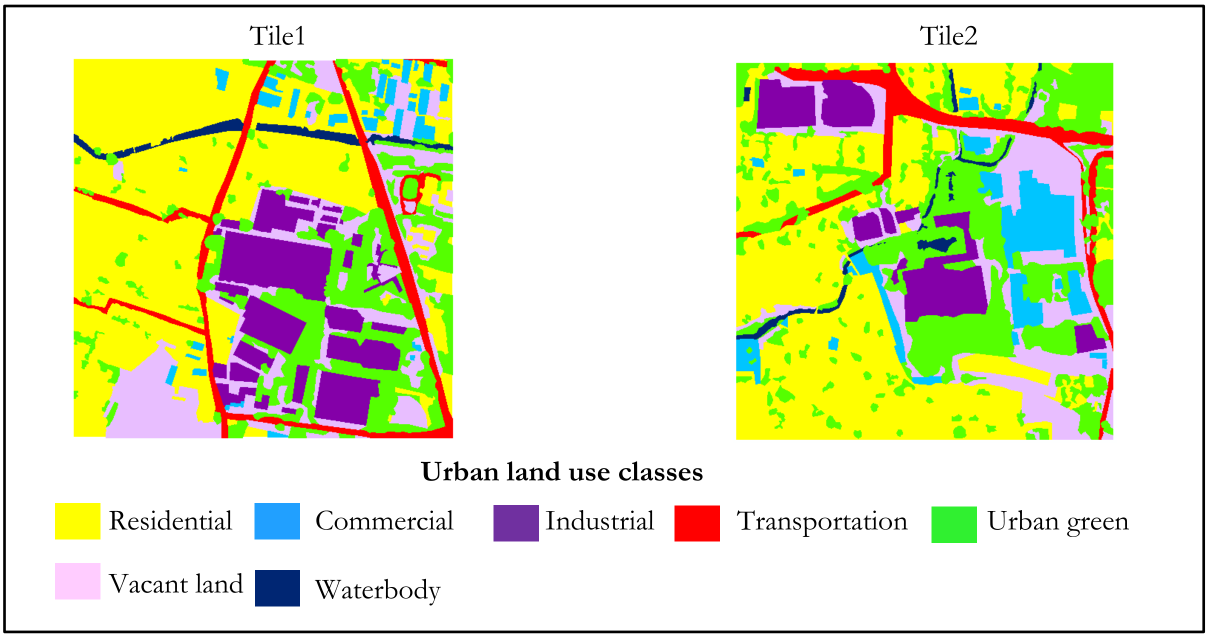


Figure 14. Shows referenced urban land use maps of Bengaluru city, 2015.

4.3. Multi-resolution image segmentation

The results of the multi-resolution segmentation employing the MRS with ESP tool on fused imagery (explained in section 3.4) is shown in Table 7, producing different scale parameters for MRS level 1, level 2, and level-3 based on the selected compactness and shape parameter. In segmentation level 1, over-segmentation is mostly observed which is shown in figure 15 which explains the over-segmentation effects of MRS. In general, over-segmentation helps to distinguish objects with a small change of compactness and shape of the object, and over-segmented objects are easy to merge by the classifiers.

Table 7. Scale parameter, shape and compactness used for the study.

Multi-resolution segmentation	Scale	Compactness	Shape	Number of segments	Mean segment size (sq. metre)
Level 1	133	0.80	0.50	3171	228.03
Level 2	223			1100	657.34
Level 3	323			508	1423.38

The MRS level 2 and level 3 shows the number of segments gradually decreasing with the increases of average segment size as compare to the level 1 (see table 7). This scenario produces the risk of under-segmentation in MRS which is shown in figure 15. The highlighted segment of MRS level 2 covers industrial and vacant land in one segment and does not fit with only the industrial land use. Similarly, MRS level 3 includes three land uses such as industrial, vacant and transportation in one segment (see figure 15). The under-segmentation is a big problem in image classification as compare to the over-segment because classifier unable to split the under-segmented object into target land use classes. The overall result shows that scale issues one of the significant problems in MRS which suggest to select an appropriate scale for classifying urban land use/cover. Therefore, selection of best MRS level is very difficult only based on visual inspection. In this regard, different MRS levels has statistically validated in feature selection and parameter tuning process (see section 4.7 and 4.8) to select best MRS level for final land use classification.

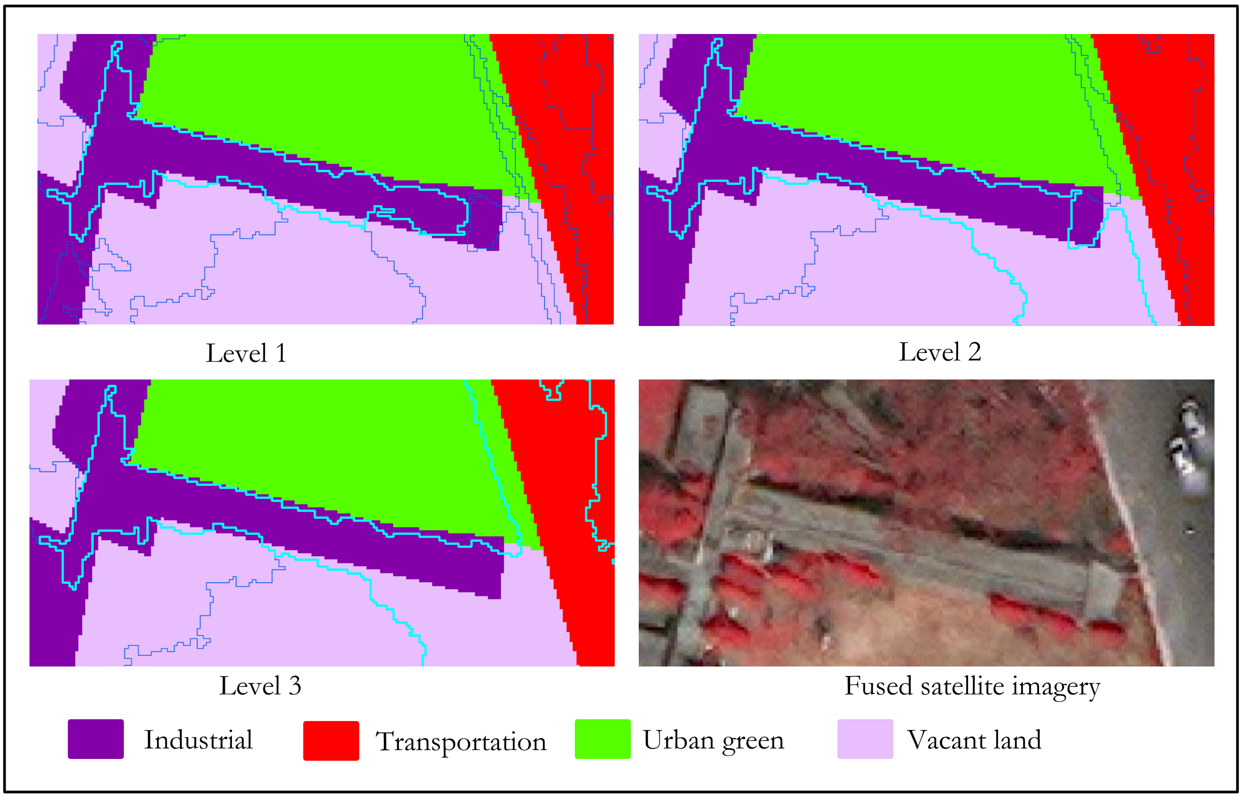


Figure 15. Shows different MRS levels compare with referenced land use and VHR imagery.

4.4. Generated blocks

This section shows the different block sizes (see figure 16) obtains by employing the method explained in section 3.5. Figure 16 shows number of the block in different block size (e.g., 29×29, 43×43, etc.) gradually decreases because of increases of scale size which shows that pure blocks(e.g., block fully covers the land use) are dramatically reduced. Consequently, heterogeneous blocks are increased. The overall result shows that scale issue one of the significant problems in BBIA which need to consider for better land use classification. Similar to the MRS levels, it is also difficult to select best block size only based on visual interpretation. In this regard, different blocks have statistically validated in feature selection and parameter tuning process (see section 4.7 and 4.8) to select best block size for final land use classification.

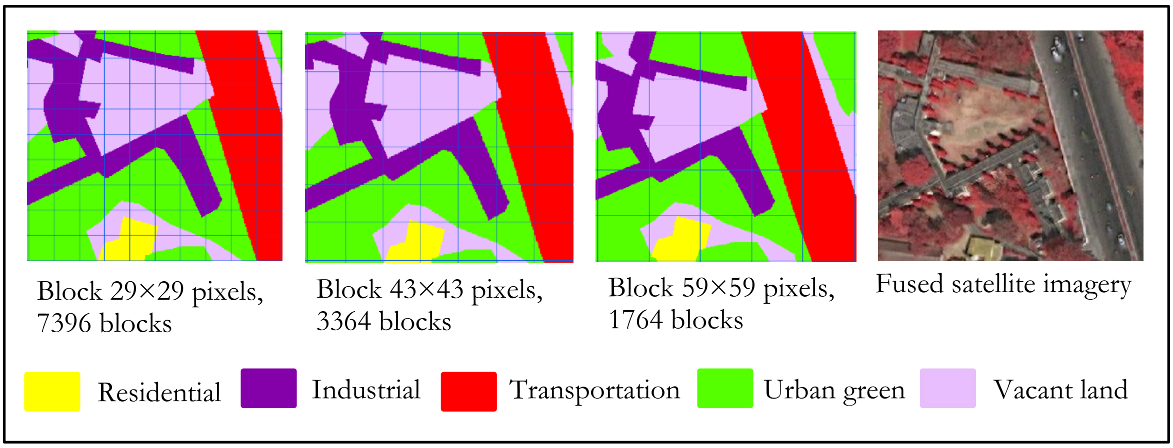


Figure 16. Shows different block size compare with the referenced land use and VHR imagery.

4.5. Extratced, aggregated and normalized Image features

In total, 135 image features (see figure 17; appendix table C2.1) have been extracted using the selected advanced methods explained in section 3.6. In OBIA-based urban land cover classification, 120 features are extracted while 125 features extracted for OBIA-Based urban land use classification. For BBIA-based urban land use classification, 121 features are extracted (see figure 17; appendix table C2.1). These image features have been normalized using the max-min method explained in section 3.7. These aggregated and normalized features have been used for selecting the best features for urban land cover/use classification.

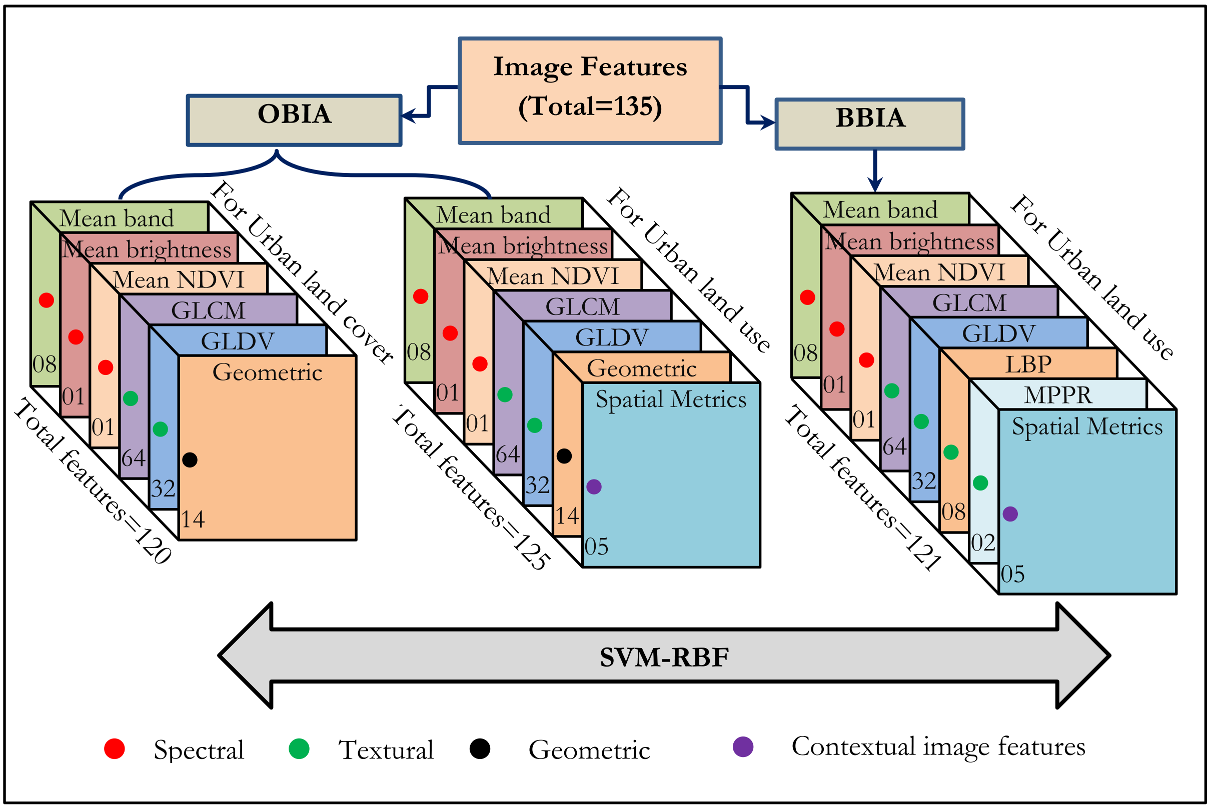


Figure 17. Shows image features for OBIA and BBIA-based urban land cover/use classification using SVM-RBF.

4.6. Training and test samples

This section shows the selected training, and test samples for urban land cover/use classification based on the strategy explained in section 3.9. Table 8 and 9 shows the training and test samples have been selected for OBIA and BBIA-based urban land use classification using SVM-RBF. In CNN-BBIA, 200 original training samples have been artificially increased to 1000, 2000 and 3000 (e.g., augmented training samples) by rotation in which training samples under each land use classes proportionally increases (e.g., each class increase by 5, 10 and 15 times). The test samples for CNN is also same as SVM-RBF. These training samples have been randomly split into training and validation samples using the 60:40 percent rule during feature selection and parameter tuning. The training and test samples at MRS level 1 for SVM-RBF-OBIA-based urban land cover classification are shown in appendix table C3.1.1, and C3.1.2. Training and test samples at MRS level 2 and 3 for SVM-RBF-OBIA-based urban land use classification are shown in appendix table C3.2.1, C3.2.2, and C3.2.3. In addition, training and test samples for block 43×43 pixel and block 59×59 pixel for SVM-RBF-BBIA-based urban land use classification are shown in table appendix C3.3.1, C3.3.2, and C3.3.3.

Table 8. Training and test samples for SVM-RBF-OBIA-based urban land use classification (MRS level 1).

Sl. no.	Training class	Tile1 (sampled domain)		Tile2 (unsampled domain)
		Number of training objects	Number of test pixels	Number of test pixels
1	Residential	50	2717828	2436234
2	Commercial	15	184944	435080
3	Industrial	30	845305	500283
4	Transportation	20	359030	340960
5	Urban green	35	1099990	1588720
6	Vacant land	20	873329	811553
7	Waterbody/Shadow	30	174575	142171
Total		200	6255001	6255001

Table 9. Training and test samples for SVM-RBF/CNN-BBIA-based urban land use classification (block 29×29).

Sl. no.	Training class	Tile1 (sampled domain)		Tile2 (unsampled domain)
		Number of training objects	Number of test pixels	Number of test pixels
1	Residential	55	2717828	2436234
2	Commercial	17	184944	435080
3	Industrial	35	845305	500283
4	Transportation	22	359030	340960
5	Urban green	36	1099990	1588720
6	Vacant land	20	873329	811553
7	Waterbody/shadow	15	174575	142171
Total		200	6255001	6255001

However, it is observed from the table 8, 9 and appendix C3.2 and C3.3 is that the variation in the number of training samples in each land use class has occurred due to increase of scale size and decrease of pure segments and blocks in each land use class.

4.7. Selected best image features

This section shows the selected best image features set for urban land cover/use classification employing the method explained in section 3.11. Out of 120 features, 25 best features have been selected for SVM-RBF-OBIA-based urban land cover classification in tile one. This selected best feature set at MRS level 1obtains highest overall accuracy (77.98%) as compare to the other features sets (see appendix table C4.1) due to having better class separability (HSIC=0.0397) as compare to others. Figure 18 and table 10 shows the best features selected from the 125 and 121 image features in tile one for SVM-RBF-OBIA and SVM-RBF-BBIA-based urban land use classification. The selected best feature set at MRS level 1 provides the highest overall accuracy (69.53 %) as compare to other features sets due to having higher class separability (HSIC=0.0586) as compare to others features set. However, selected best feature set at MRS level 1 provides the highest overall accuracy (69.53 %) as compare to the feature set at MRS level 2 and 3 (see figure 18 and appendix table C4.2) due to the increasing effects of under-segmentation (see section 4.3). Similarly, selected best image feature set at block 29×29 obtains highest overall accuracy (67.70%) as compare to the other feature set due to having higher class separability (HSIC = 0.0574). However, selected best feature set at block 29×29 provides highest overall accuracy (67.70%) as compare to the feature set at block 43×43 and block 59×59 due to increasing effects of the mixed block caused by scale issue (see section 4.4) (see figure 18 and appendix table C4.3). The best features which have been selected from tile one and same features have been used for tile two for domain adaptation.

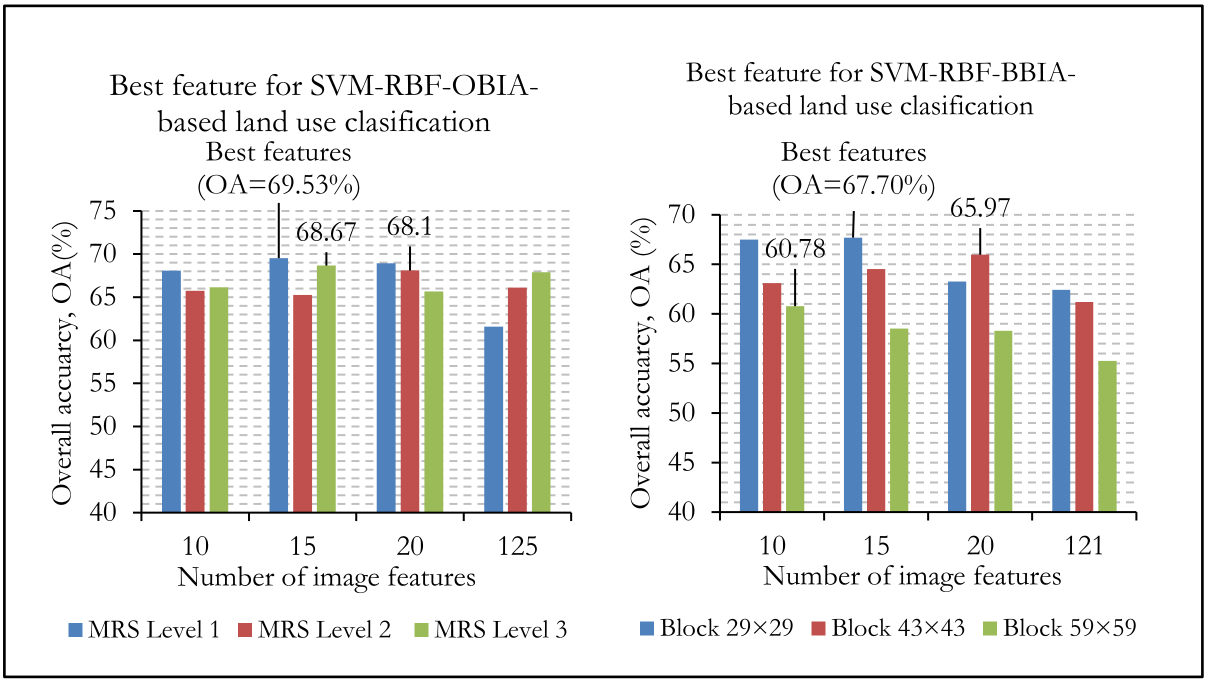


Figure 18. Shows selection of best image features for SVM-RBF-OBIA and SVM-RBF-BBIA-based land use classification using SFS-HSIC.

Table 10. Proposed best fifteen images features for urban land use classification using SFS-HSIC.

Number best of Features	SVM-RBF-OBIA (MRS level 1)		SVM-RBF-BBIA (block 29×29)	
	Features type	Name of the best features	Features type	Name of the best features
1	Spectral	Meanband3	Spectral	Meanband1
2		Meanband4		Meanband2
3		Meanband5		Meanband3
4		Meanband6		Meanband4
5		Meanband7		Meanband5
6		Meanband8		Meanband6
7		Mean NDVI		Meanband7
8	Textural	GLCMcorrelation band1		Meanband8
9		GLDVentropy band1		Mean brightness
10		GLDVentropy band2		Mean NDVI
11		GLDVentropy band8	Textural	GLCM entropy band8
12	Contextual	Aggregation index		GLDV entropy band7
13		Fractal dimension	Contextual	Aggregation index
14		Cohesion		Cohesion
15		Largest patch index		Largest patch index

The above result shows that features selection one of the very important consideration to select the appropriate features which having better class separability for improving the urban land cover/use classification accuracy. It is also observed that spectral features along with textural and contextual features are very important for classifying urban land use because urban land use is mostly separated in terms of texture and contextual features. In addition, one of the important findings is that most of the cases GLCM widely used for classifying urban land use (e.g., Kuffer et al. 2017; Herold et al. 2003; Yanchen et al. 2014) but result shows GLCM along with GLDV in most important for addressing texture in better way using VHR imagery for urban land use classification (Aguilar et al., 2012). Another important obesrvation is that

LBP, MPPR and geometric features are not selected perhaps due to combine effects of robust textural features such as GLCM, GLDV and contextual fetaures such as spatial metrics.

However, in feature selection process, best gamma and fixed cost are equal to 100 are used for SVM-RBF. However, the accuracy of the best feature sets in different MRS levels and block levels might be varied while employing the best gamma and best cost value together. Thus, best feature set with higher overall accuracy at MRS level 1 (15 features), MRS level 2 (20 features) and MRS level 3 (15 features) as well as block 29×29 (15 features), 43X43 (20 features) and block 59×59 (10 features) (see figure 18, appendix table C4.2 and C4.3) have been used for parameter tuning using k-fold cross-validation. This helps to select the best image feature set at best MRS level and block level for urban land use classification in a very efficient manner.

4.8. Selcted best parameters

This section shows the selected best parameters of SVM-RBF and CNN for urban land cover/use classification employing the method explained in section 3.12. which are explained below:

4.8.1. Best parameters of SVM-RBF

This section shows the results obtains from the holdout grid search 10-fold cross-validation explained in section 3.12.1. The result shows that 25 best features at MRS level 1 shows that the best gamma (0.1668) and best cost (1000) provides highest overall accuracy (78.21 %) as compare to others set of parameters for SVM-RBF-OBIA-based urban land cover classification (see appendix table C5.1.1). Table 11 shows the best parameters at different best features at different MRS and block level for SVM-RBF-OBIA and SVM-RBF-BBIA -based urban land use classification. The result shows that the overall accuracy of different best features at different MRS and block level has improved while employing best parameters as compare to best gamma with fixed cost as explained in section 4.7. Therefore, it is also observed that best features at MRS level 1 still provides highest overall accuracy (70.58%) as compare to the best features at MRS level 2 MRS level 3 while employing best parameters (see appendix table C5.1.3, table 11). Similar, outcome also shows in different block level (see appendix table C5.1.4, table 11). Therefore, the overall result shows that parameter tuning one of the foremost consideration to improve the overall classification accuracy by penalizing the cost of overfitting of the classifier (e.g., SVM-RBF). The results also prove that the effect of under-segmentation in MRS level 2 and 3, as well as the effect of the mixed block in block size 43×43 and 59×59, still exits while employing best parameters. Therefore, best parameters, best features at MRS level 1 (best MRS level) and block 29×29 (best block size) are selected in tile one (see highlighted column of table 11), and same has used for tile two for domain adaptation.

Table 11. Parameter tuning for SVM-RBF-OBIA and SVM-RBF-BBIA using 10-fold cross-validation.

SVM-RBF parameters	SVM-RBF-OBIA			SVM-RBF-BBIA		
	MRS level 1 15 features	MRS level 2 20 features	MRS level 3 15 features	**Block 29×29 15 features**	Block 43×43 20 features	Block 59×59 10 features
Best gamma	**0.0599**	0.1668	0.4642	**1.292**	0.4642	1.292
Best cost	**1000**	129.155	359.381	**16.681**	16.681	16.681
Overall accuracy	**70.58**	68.10	69.08	**68.37**	67.46	60.92

- **Experiment of training samples size on proposed SVM-RBF best parameters**

The experiment on the size of training samples is essential because the accuracy of the classifiers is affected by the size of training samples due to the effect of Hughes phenomena. Thus, different size of training samples has been experimented using the best parameters of SVM-RBF. The proposed best parameter of SVM-RBF-OBIA for urban land cover classification provides highest overall classification (78.21%) on 200 training samples as compare to the other sample sizes (see appendix, table C5.1.2). Table 12 shows that 150 training samples provide highest overall classification accuracy as compare to other sizes of training samples on proposed best parameters of SVM-RBF-OBIA (70.83%) and SVM-RBF-BBIA (68.51%) for urban land use classification.

Table 12. Experiment of size of training samples for urban land use classification.

Size of training samples	Overall accuracy	
	SVM-RBF-OBIA	SVM-RBF-BBIA
50	65.49	60.93
100	66.19	63.39
150	**70.83**	**68.51**
200	70.58	68.37

The result shows that the absence of optimum size of training samples has a serious effect on overall classification accuracy. Therefore, the result also proves that experiment of different training samples size is a good choice to improve the overall classification accuracy by reducing the effect of Hughes phenomena (Mboga, 2017). Therefore, this proposed best training samples size (150) of tile one which has been used for urban land use classification of tile two.

4.8.2. Best parameters of CNN

This section provides the best hyperparameters of the CNN which have been obtained employing the parameters tuning strategy explained the section 3.12.2. This has done using 1000 training samples and 6,255,001 test samples with stochastic gradient descent. The best learning and regularization parameters and CNN hyperparameters are explained below:

- **Learning and regularization parameters**

The best learning and regularization parameters of the CNN-BBIA-based urban land use classification have been obtained using 6-fold cross-validation (see appendix table C5.2.1.1, C5.2.1.2, and C5.2.1.3, figure C5.2.1.1.). The network obtains 65.12 % overall classification (see appendix table C5.2.1.3) on best learning and regularization parameters (see table 13) with the fixed CNN configuration (see table 14). The overall accuracy has decreased for other learning and regularization parameters due to overfitting of the network. In the learning and regularization parameters tuning phase, zero padding is 2 and stride 1 have been used for convolutional layers while zero padding 1 and stride 2 used for the pooling layers. In addition, stride 1 and no zero padding have been used for FC layer.

Table 13. Best learning and regularization parameters of CNN-BBIA-based urban land use classification.

Hyperparameters	Value
Batch size	10
Maximum number of epoch	1000
Momentum	0.90
Learning rate	0.001
Weight decay	0.01
Dropout rate in (D1 & D2)	(0.25, 0.5)

Table 14. Fixed CNN configuration used for selecting best learning and regularization parameters.

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in C_{1-2} and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolution layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

The result shows that the use of inappropriate learning and regularization parameters has a serious effect on overfitting of the network which affects the overall classification accuracy. Therefore, tuning of learning and regularization parameters is one of the important consideration in CNN for urban land use/cover classification (Bergado, 2016; Mboga, 2017).

- **CNN hyperparameters**

The 4-fold cross-validation of the CNN hyperparameters such as the number of filters, size of filters, number of convolutional layers and number of fully-connected layers have been carried out which are explained below:

- **Experiment on number of filters**

The experiment on different number of filters have been carried out using the selected best learning and regularization parameters (see appendix table C5.2.2.1) and with fixed others CNN hyperparameters (see appendix table C5.2.2.2). Increasing the number of filters helps to learn complex features in one way and increased huge parameters in another way (Bergado, 2016; Mboga, 2017). The increases of huge parameters prone to overfitting of the network which reduces the overall classification accuracy (CS231n, 2018). As a result, the network provides 65.12% overall accuracy on the best 8 filters while accuracy has decreased for the other filters due to overfitting of the network (see appendix table C5.2.2.3, figure 19). The zero padding, stride for convolutional and FC layers are similar as mentioned in the tuning phase of the learning and regularization parameters.

- **Filter size experiment**

The filter size experiment has been carried using best learning and regularization parameters (see appendix table C5.2.2.4) and with fixed others CNN hyperparameters (see appendix table C5.2.2.5). The increase of filter size learns large and more complex spatial pattern (e.g., edge and gradient) to address the particular land use classes in one way and also to increase the parameters in another way (Bergado, 2016). Too large filter overestimates the spatial pattern for a particular land use classes as well as large parameters leads to overfitting of the network. The overfitting of the network reduces the overall classification accuracy. As a result, the network provides the highest overall accuracy (65.12%) on filter size 5X5 pixels as compare to the other filter sizes (see appendix table C5.2.2.6, figure 19). The zero padding for filter size 5, 15 and 25 are 2, 7 and 12 while stride 1 have been used for convolutional layers. The stride 1 and no zero padding have been used for FC layer.

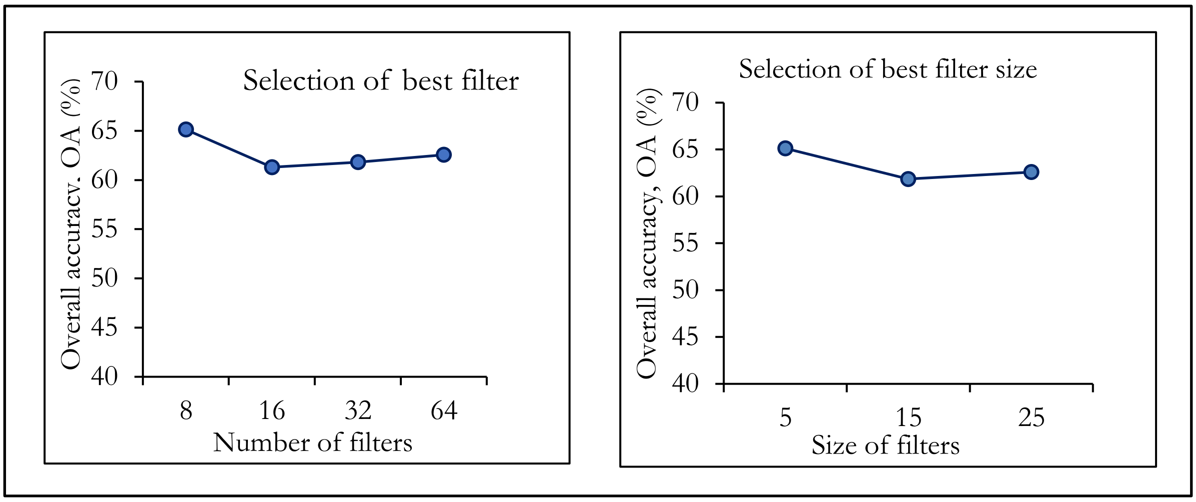


Figure 19. Shows overall accuracy varied with different number of filters and filter sizes.

- **Experiment of different convolutional layers**

The experiment of the different convolutional layer (C_n) with the fixed FC layers (e.g., FC=1) which leads to understand how the depth of the network helps to improve the classification accuracy. In addition, increasing the number of convolutional layers sometimes provides more abstract features to improve the classification accuracy (Mboga, 2017). However, increasing the number of convolutional layers also sometimes oversimplify the features due to frequent drop of parameters thorough max-pooling in each convolutional layer. As a result, two convolutional layers (C_2) with the fixed FC layers one with the best learning and regularization parameters (see appendix table C5.2.2.7) and the best CNN hyperparameters (see appendix table C5.2.2.8) provides highest overall classification accuracy (65.12%) as compare to other convolution layers (see appendix table C5.2.2.9; figure 20). The zero padding, stride for convolutional and FC layers are similar as mentioned in the tuning phase of the learning and regularization parameters and number of filters experiment.

- **Experiment of different fully-connected layers**

The experiment of different FC layers has been carried out with the fixed convolutional layer (e.g., 2) and best learning and regularization parameters (see appendix table C5.2.2.10) and best CNN hyperparameters (see appendix table C5.2.2.11). The FC layer is termed as the dense layer of CNN.

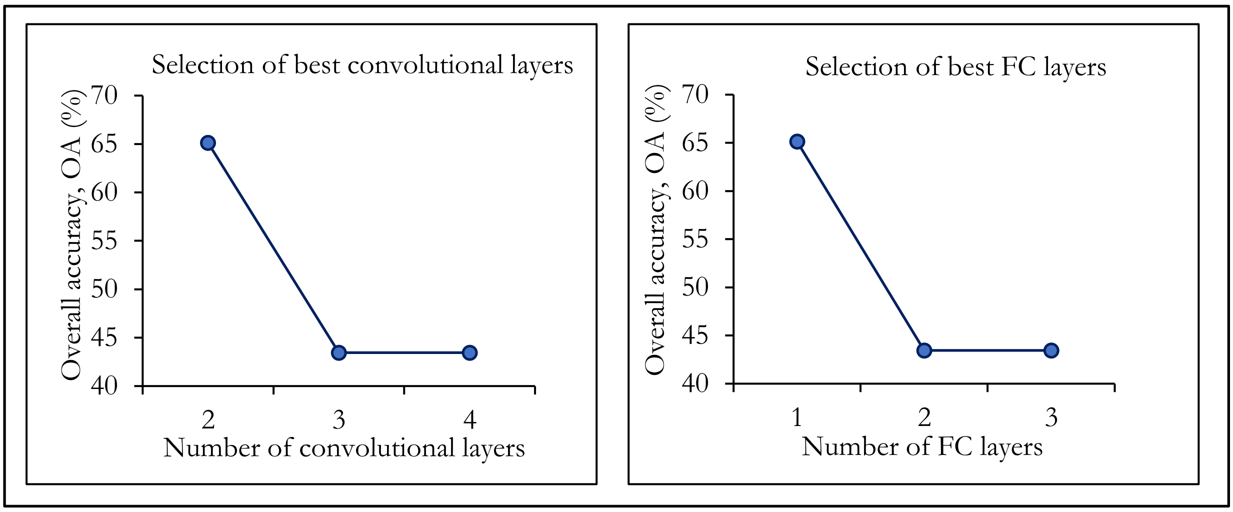


Figure 20. Shows overall accuracy varied with different number of convolutional and FC layers.

When increasing the FC layers, the network becomes complex. A complex network increases the risk of overfitting of the network. As mentioned above overfitting the network reduces the classification accuracy. The result shows that FC layer one gets the highest overall accuracy (65.12%) as compare to FC layer two and three (see appendix table C5.2.2.12, figure 20). The zero padding, stride for convolutional and FC layers are similar as mentioned in the tuning phase of the learning and regularization parameters, number of filters experiment and number of convolutional layers experiment. The tuning of CNN hyperparameters shows that overall classification is affected by the overfitting of the network while employing unsuitable hyperparameters. Therefore, along with learning and regularization parameters, tuning of CNN hyperparameters is also essential considering to learn best local contextual features to improve the overall classification accuracy (Bergado, 2016 and Mboga, 2017). The best CNN parameters as follows:

- **Proposed best parameters and architecture of CNN**

Based on the 10-fold cross-validation of CNN parameters, the proposed learning, and regularization parameter, CNN hyperparameters and architecture are shown in table 15 and 16. The proposed CNN architecture obtains two convolutional layers and one FC layer. The padding is 2 and stride 1 have been used for convolutional layers while zero padding 1 and stride 2 used for the pooling layers. In addition, stride 1 and no zero padding have been used for FC layer. The proposed CNN architecture has been developed on tile one, which provides 65.12% overall classification accuracy on 1000 training samples and 6,255,001 test samples. This proposed CNN architecture is applied in tile two for domain adaptation.

Table 15. Proposed best learning and regularization parameters.

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.001
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0.25, 0.5)

Table 16. Proposed best CNN configuration and hyperparameters.

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in C_{1-2} and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

- **Experiment of training sample sizes with proposed best parameters and architecture of CNN**

Table 17 shows that the proposed best learning and regularization parameter (see table 15), CNN hyperparameters and configuration (see table 16) provides the highest overall classification accuracy on 1000 training samples (augmented training samples) as compare to the other training samples. Overall classification accuracy on training samples 200 is low (55%) because the CNN performs better on higher samples size or augmented training samples. However, too high samples size for a specific CNN architecture, network gets overfitted and consequently reduces the overall classification accuracy (see table 17). The result shows that use of the inappropriate size of training samples has a serious effect on overfitting of the network because of the effects of Hughes phenomena. In addition, the result also shows that use of augmented training samples one of the important choice to reduce the overfitting of the network (Mboga, 2017). Therefore, the result proves that experiment on different size of training samples as well as on augmented training samples has improved the overall classification accuracy by reducing the overfitting of the network (see table 17).

Table 17. Experiment of size of training samples for urban land use classification.

Size of training samples	Overall accuracy
200	55.00
1000	**65.12**
2000	62.77
3000	62.32

4.9. Urban land cover/use classification and accuarcy assesement

This section presents the urban land cover/use classification, validation, and accuracy. The urban use classification has carried out in sampled (tile one) and unsampled (tile two) domain which are explained below:

4.9.1. SVM-RBF-OBIA-based urban land cover classification

Based on the proposed best features set, SVM-RBF parameters and training samples size (see section 4.7 and 4.8.1), the SVM-RBF-OBIA provides 78.21% overall classification accuracy and kappa 0.6507 for urban land cover classification in tile one (see appendix figure C6.1, table C6.1). In addition, while adopting the similar proposed parameters as used in tile one, the SVM-RBF-OBIA provides 75.52 % overall classification accuracy and kappa 0.6182 for urban land cover classification in tile two (see appendix figure C6.2, table C6.2). These classified urban land cover have been used for the contextual features (e.g., spatial metrics) extraction for urban land use classification (see section 3.6).

4.9.2. SVM-RBF-OBIA-based urban land use classification

Table 18 shows the accuracy of SVM-RBF-OBIA-based urban land use classification obtains using proposed best features, SVM-RBF parameters and best size of training samples explained in section 4.7 and 4.8.1. Table 18 shows the overall accuracy is 65.30 % and kappa is 0.5467 (average of tile 1 and 2) in SVM-RBF-OBIA-based urban land use classification. In addition, table 18 shows that both user and producer accuracy of commercial land use is low as compare to residential and industrial land use in tile one which explains that commercial land use mostly mixed with residential, industrial and vacant land use (see appendix table C7.1, C7.2 and figure 21). In addition, results also show that user accuracy of waterbody/shadow is low while producer accuracy is high. This explains that most of the land use misclassified into waterbody/shadow because such land uses have shadow effects while waterbody/shadow is less misclassified into other land use classes. The result also shows that accuracy is significantly dropped in tile two due to very significant drop in the accuracy of the commercial, industrial, vacant land and waterbody/shadow. Therefore, the overall result (both tiles) shows that complexity of commercial, industrial and vacant land use as well as shadow effect one of the cause for reducing the overall accuracy.

Table 18. Accuracy of SVM-RBF-OBIA-based Urban land use classification.

Land use classes	Accuracy (%)			
	Tile one (sampled domain)		Tile two (unsampled domain)	
	User	Producer	User	Producer
Residential	79.97	75.98	66.98	72.42
Commercial	39.41	16.90	17.47	16.61
Industrial	77.00	60.49	16.57	8.71
Transportation	53.50	63.24	46.00	75.04
Urban green	75.77	84.49	85.53	79.66
Vacant land	58.02	59.12	51.59	38.38
Waterbody/shadow	38.57	85.86	19.29	41.29
Overall	70.83		60.30	
Kappa	0.6231		0.4703	

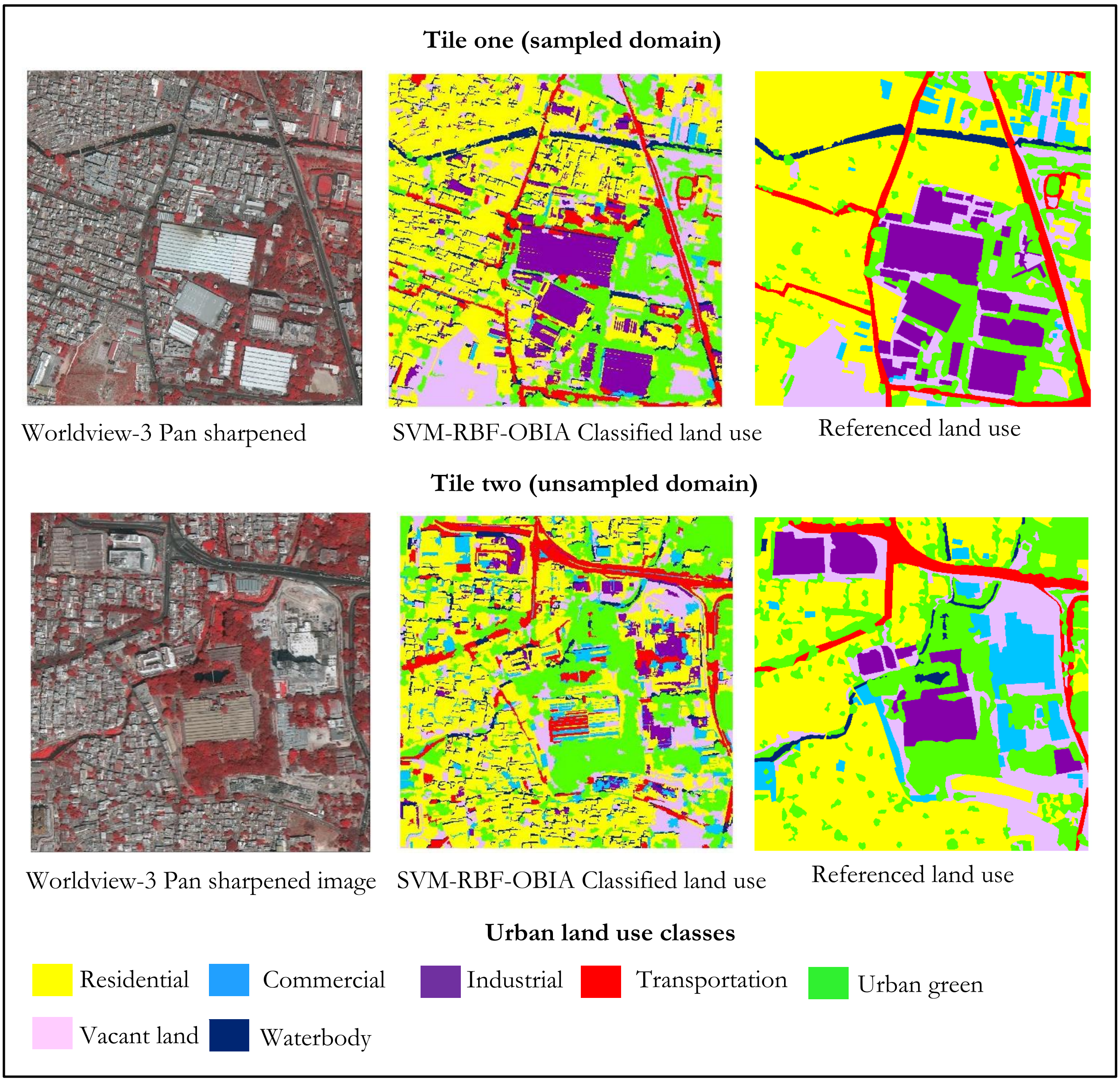


Figure 21. Shows SVM-RBF-OBIA-based urban land use classification of tile one and tile two.

4.9.3. SVM-RBF-BBIA-based urban land use classification

Table 19 shows that overall accuracy (OA) is 56.64% and kappa is 0.3900 (average of tile 1 and 2) of SVM-RBF-BBIA-based urban land use classification which is less (overall accuracy -8.66% and kappa -0.1567) than the SVM-RBF-OBIA because misclassification rate is high in most of the land use classes such as commercial, industrial, vacant land and waterbody/shadow (see appendix table C8.1, C8.2 and figure 22). The one of the very important observation is that land use classes with a linear shape (waterbody, transportation) misclassified within or between the land use classes because regular grid unable to extract the complete shape of linear land uses as compare to OBIA. Another important observation is that this classification minimizes the effect of shadow by merging the shadow with other land use classes.

Table 19. Accuracy of SVM-RBF-BBIA-based urban land use classification.

Land use classes	Accuracy (%)			
	Tile one (sampled domain)		Tile two (unsampled domain)	
	User	Producer	User	Producer
Residential	71.99	87.66	50.17	83.89
Commercial	32.68	25.18	33.83	5.91
Industrial	78.24	51.70	27.66	2.76
Transportation	36.31	61.41	22.55	54.92
Urban green	83.89	64.50	93.70	18.50
Vacant land	63.91	48.91	24.05	23.36
Waterbody/ shadow	50.62	35.59	37.69	32.75
Overall accuracy	68.51		44.77	
Kappa	0.5600		0.2200	

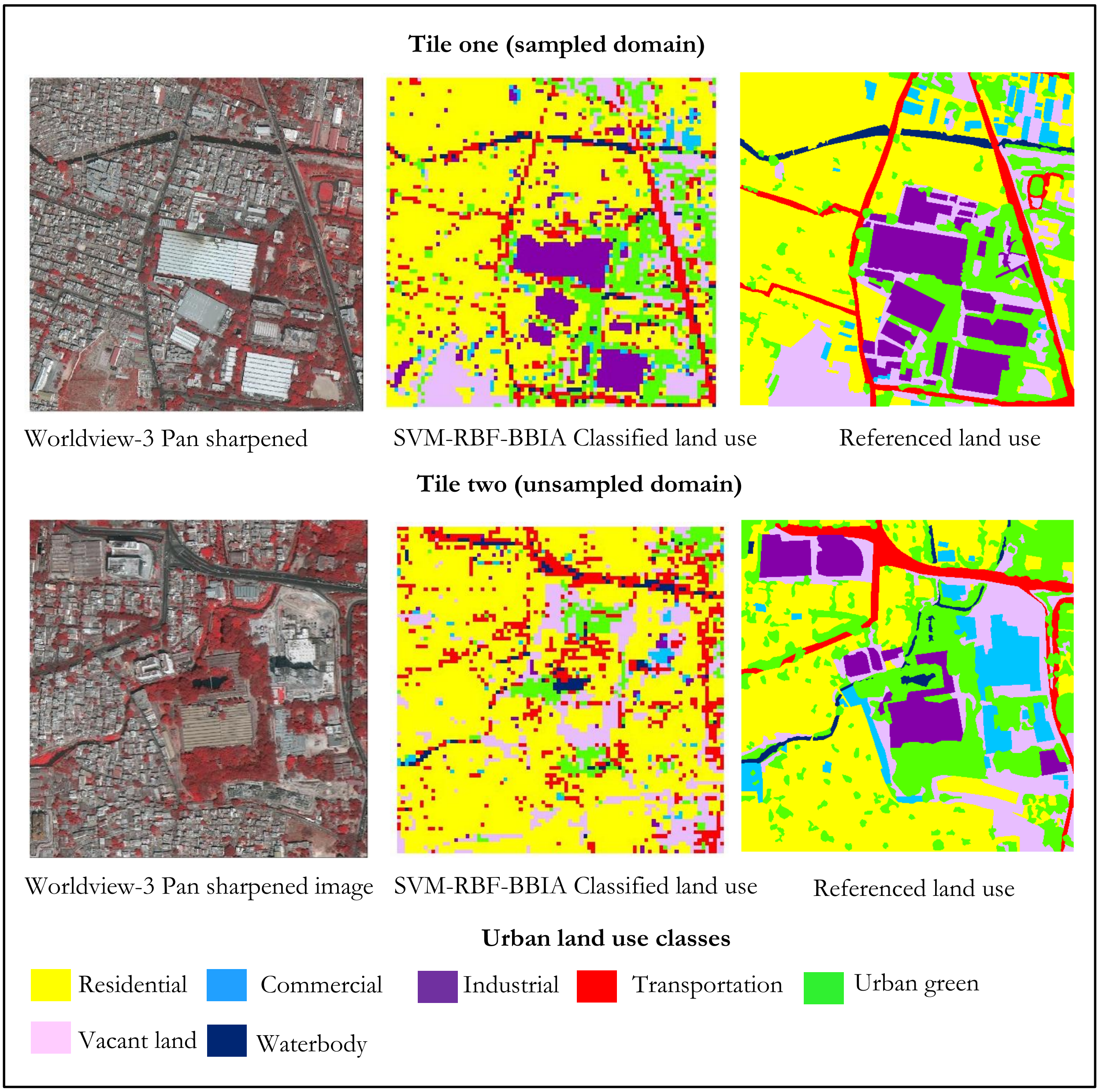


Figure 22. Shows SVM-RBF-BBIA-based urban land use classification of tile one and tile two.

4.9.4. CNN-BBIA-based urban land use classification

Table 20 shows that CNN-BBIA obtains overall accuracy is 57.51%, and kappa is 0.4245 (average of tile 1 and 2) which is higher (overall accuracy +0.87%, +0.0345) than the SVM-RBF-BBIA and lower (overall accuracy -7.79%, kappa -0.1222) than the SVM-RBF-OBIA. This shows that the misclassification accuracy of commercial, industrial, vacant land, and waterbody/shadow (see figure 23 and appendix table C9.1 and 9.2) quite higher than SVM-RBF-OBIA while less than the SVM-RBF-BBIA. One of the very important observation is that land use classes with a linear shape (e.g., transportation, waterbody) has a good classification in both sampled and unsampled domain as compare to the SVM-RBF-BBIA because of the better edge detection ability of the CNN. In addition, another important observation is that this classification minimizes the effect of shadow by merging with other land use classes similar to the SVM-RBF-BBIA. This is because block/patch bigger than the shadows is automatically merged with other land use classes.

Table 20. Accuracy of CNN-BBIA-based urban land use classification.

Land use classes	Accuracy (%)			
	Tile one (sampled domain)		Tile two (unsampled domain)	
	User	Producer	User	Producer
Residential	72.35	73.55	61.51	80.31
Commercial	20.19	27.59	6.98	4.65
Industrial	73.65	48.44	13.62	11.54
Transportation	44.58	62.63	54.50	66.34
Urban green	85.19	73.62	98.31	35.98
Vacant land	46.37	59.61	18.77	30.28
Waterbody/ shadow	55.65	33.77	78.07	29.74
Overall	65.12		49.89	
Kappa	0.5274		0.3216	

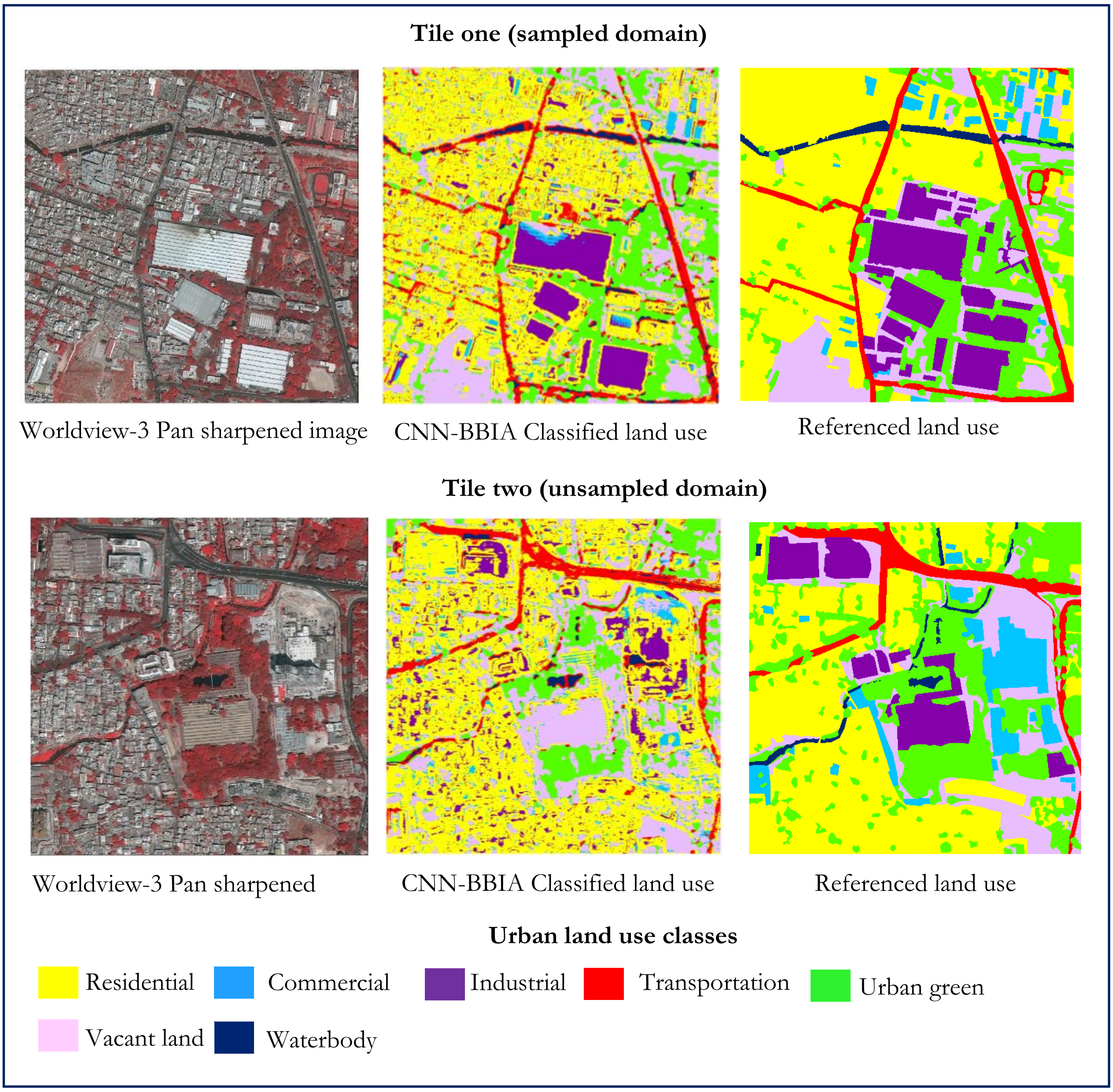


Figure 23. Shows CNN-BBIA-based urban land use classification of tile one and tile two.

Form the above results, it is observed that misclassification of commercial, industrial is mostly happened due to sharing of common features information with the residential and vacant land as well as between commercial, industrial. This occurs due to increase of mixed land use pattern in the selected area. In addition, another important observation is that overall accuracy, as well as the accuracy of other land uses is affected by sharing of common features information of the waterbody and shadow because most of the land uses have shadow effect in VHR imagery. It is also observed that effect of shadow is mostly minimized in SVM-RBF-BBIA and CNN-BBIA while this effect significant in SVM-RBF-OBIA because the segmentation follows the shape of the object. The overall results show that SVM-RBF-OBIA provides higher classification accuracy as compare to SVM-RBF-BBIA and CNN-BBIA while CNN-BBIA provides higher classification accuracy as compare to SVM-RBF-BBIA. In addition, very important observation is that the accuracy is affected by the combination of classification algorithms and classification approach which is most important consideration in land use classification. For detailed explanation needs more accuracy assessment indices which are explained in section 4.10.

4.10. Performance evaluation

In this section, comparison of accuracy assessment indices of the selected machine learning algorithms has been discussed in sampled and unsampled domain employing methods explained in section 3.15. This helps to assess the performance and robustness of the classifiers for urban land use classification which are explained below:

4.10.1. Performance based on quantitative indices

Table 21 shows that F1-score of most of the urban land uses such as residential, commercial, industrial, transportation in OBIA outperforms the BBIA. This explains that urban land uses are highly associated with the images features related to shape, size and orientation of the land use objects along with others features. In addition, F1-score of waterbody/shadow in BBIA outperforms the OBIA which explains that the OBIA is suffering from the effect of shadow in VHR imagery. The result also shows that F1-score of most of the urban land uses in CNN outperforms the SVM-RBF. This is because classifying the complex urban land use is highly related to the more complex contextual features learned by CNN.

Table 21. Performance of selected machine learning algorithms for urban land use classification (compiled tile one (sampled domain) and tile two (unsampled domain) see appendix table C10.1, C10.2).

Land use classes	SVM-RBF-OBIA			SVM- RBF-BBIA			CNN-BBIA		
	Recall	Precision	F1-score	Recall	Precision	F1-score	Recall	Precision	F1-score
Residen.	73.48	74.20	73.76	61.08	85.78	70.92	66.93	76.93	71.30
Commer.	28.44	16.76	20.35	33.26	15.55	19.26	13.59	16.12	14.45
Industri.	46.79	34.60	39.59	52.95	27.23	33.64	43.64	29.99	35.47
Trans.	49.75	69.14	57.50	29.43	58.17	38.81	49.54	64.49	55.97
Ur.green.	80.65	82.08	81.19	88.80	41.50	51.91	91.75	54.80	65.83
Vac.land	54.81	48.75	51.29	43.98	36.14	39.56	32.57	44.95	37.67
Water/ shadow	28.93	63.58	39.77	44.16	34.17	38.42	66.86	31.76	42.55
Overall	51.84	55.59	51.92	50.52	42.65	41.79	52.13	45.58	46.18

Table 22. Performance of selected machine learning algorithms based on quantitative accuracy.

Tile	SVM-RBF- OBIA		SVM- RBF-BBIA		CNN-BBIA	
	Overall accuracy	Kappa	Overall accuracy	Kappa	Overall accuracy	Kappa
1 -sampled	70.83	0.6231	68.51	0.5600	65.12	0.5274
2-unsampled	60.30	0.4703	44.77	0.2200	49.89	0.3216
Overall	65.57	0.5467	56.64	0.3900	57.51	0.4245

Therefore, the overall F1-score, overall accuracy, and overall kappa (see table 21 and 22) shows that OBIA outperform the BBIA and similarly CNN outperform the SVM-RBF. Another important observation also is that OBIA outperforms the CNN. The similar results also observed in the unsampled domain (see table 21 and 22). This is because geometrical characteristics (e.g., shape, size, orientation, etc.) of OBIA is an added advantage over others robust spectral, textural (e.g., GLCM, GLDV) and contextual characteristics (e.g., spatial metrics) of land use classes for classifying complex urban land use as compare to BBIA and CNN. Therefore, in previous research proven that CNN is outperform the SVM-RBF and also proven that feeding of robust handicraft features (e.g., GLCM or CNN learned features) in SVM-RBF is most of the cases competitive with the CNN (e.g., Mboga, 2017). Therefore, overall results also show that OBIA are more transferable and robust as compare to BBIA and similarly CNN is more transferable and robust as compare to SVM-RBF.

4.10.2. Performance based on locational indices

In terms of the locational agreement, table 23 shows that the overall Klocation (e.g., kappa for location) and Kno (overall kappa both in quantity and location) in OBIA outperforms the BBIA and similarly, the overall kappa (Kno) in CNN outperforms the SVM-RBF. The similar results also observe in the unsampled domain (tile two).

Table 23. Performance of selected machine learning algorithms based on locational accuracy.

Tile	SVM-RBF- OBIA		SVM- RBF-BBIA		CNN-BBIA	
	Klocation	Kno	Klocation	Kno	Klocation	Kno
1 -sampled	0.6695	0.6596	0.6877	0.6326	0.5926	0.5931
2-unsampled	0.5361	0.5368	0.4230	0.3555	0.4498	0.4153
Overall	0.6028	0.5982	0.5554	0.4941	0.5212	0.5042

Therefore, in terms of the locational agreement, the result also proves that the OBIA is more robust and transferable as compare to BBIA and similarly, CNN is more robust and transferable as compare to SVM-RBF for urban land use classification.

4.10.3. Performance based on classified map quality

Visually the classified map of OBIA looks better in terms of quality as compare to the BBIA using the reference (see appendix figure C10.1) and local land use maps (see appendix figure C10.2). Similarly, visually the classified map of CNN also looks better as compare to the SVM-RBF (see appendix figure C10.1 and C10.2). This comparison of classified map quality is based on overall shape, size, and misclassification of land use classes. The misclassification visually shows the mixed and scatters pattern of urban land use. Therefore, in terms of classified map quality, it is concluded that OBIA is more relevant as compare to BBIA and similarly, CNN is more relevant as compare to SVM-RBF for urban land use classification and to address the local land use classification scheme.

Therefore, based on the quantitative, locational and qualitative performance assessment the result shows that OBIA is robust and more transferable as compare to BBIA and similarly, CNN is also robust and more transferable as compare to SVM-RBF for urban land use classification from VHR imagery. Hence, the results suggest that the use of CNN in OBIA is an optimum choice to develop a most promising urban land use classification approach from VHR imagery for developing countries like India.

4.10.4. Performance based on classification time

In this study, the selection of best image features, parameters and undergone land use classification in SVM-RBF for OBIA or BBIA takes half an hour for a single tile. In addition, based on the best parameters, the CNN-BBIA takes 5 hours to learn the features and to produce the final classified map of one single tile. However, it is also observed that overall processing time is quite high in SVM-RBF as compare to the CNN because additional time takes for preparing and exploring the handicraft image features. Therefore, in larger city scale, the CNN is more relevant as compare to SVM-RBF because of self-learning ability of CNN.

4.11. Key summary of the results and discussions

The key summary of the results and discussion is that classification of urban land use from VHR imagery is affected by the several factors such as types and number of image features, scale issues, parameters and types of classifiers and size of training samples which are very sensitive to the overall classification accuracy and quality of the results. Based on the sensitivity analysis results shows that OBIA outperforms the BBIA and similarity CNN outperforms SVM-RBF. The OBIA outperforms the BBIA because one the important reason is the geometrical characteristics of the OBIA which directly link with geometrical properties of the urban land uses. In addition, CNN outperforms the SVM-RBF because CNN extracts more complex contextual features which address the texture, edge, and gradient of the different urban land uses. In this study, post-processing (for improving accuracy) has not done because in this case need to link with other algorithms like Conditional Random Field (CRF) which require further study. The conclusion of the results and discussion has explained in section 5.

5. CONCLUSION AND RECOMMENDATION

In this section conclusions, reflections on research objectives and questions, research contribution, limitations, and recommendations are drawn for future research:

5.1. Reflection on the research objective and research question

The main objective of this study was to evaluate the performance of different machine learning algorithms for urban land use mapping. Thus, based on the research objective, a wide variety of image features (i.e., 135 spectral, textural, geometric and contextual image features) were extracted from VHR imagery, and such features were selected employing SFS-HSIC for SVM-RBF-based urban land use classification. Based on this features selection it was observed that only 15 are very efficient image features out of the total 135 features. In addition, selection of optimum parameters of SVM-RBF employing hold-out cross-validation improved the overall land use classification accuracy. However, it was also observed that varying the size of training samples affected the overall classification accuracy. In SVM-RBF, 150 (out of 200) is the best training samples size to reduce the effect of Hughes phenomena. This improved the overall urban land use classification. In OBIA-based urban land use classification, the scale issue in segmentation (using MRS) has a very serious effect on the overall classification accuracy. It was observed that the MRS level 1 (i.e., scale 133) provided the highest overall classification as compared the MRS level 2 (i.e., scale 223) and MRS level 3 (i.e., scale 323). Similarly, scale issues were also observed in BBIA-based urban land use classification in which the block size 29X29 showed the highest overall accuracy as compared to the block size 43X43 and 59X59. Thus, based on the different accuracy assessment indices, map quality, time and domain adaptation, it is concluded that SVM-RBF-OBIA is more relevant and robust as compared to the SVM-RBF-BBIA for urban land use classification from VHR imager in the Indian context.

In CNN-BBIA based urban land use classification, it was observed that consistent use of learning and regularization parameters and hyperparameters of CNN configuration reduced the risk of overfitting of the network. This is because overfitting of the network reduced the overall classification accuracy. In addition, one of the important observation from the network is that the co-adaptation of neurons was detected in both convolutional and FC layer. Consequently, the use of dropout in both layers improved the overall classification accuracy. Thus, experiments on the depth of the network showed a change of overall classification accuracy because of the variation of learned contextual image features. Therefore, one of the final observation is that the overall classification was improved while using augmented training samples instead of original training samples because data augmentation is one of the important solutions to resolve the overfitting of the network. Therefore, based on different accuracy assessment indices, map quality, time and domain adaptation, finally, it is concluded that the OBIA is more relevant and robust as compared to BBIA for urban land use classification from VHR imager in the Indian context. However, using BBIA, CNN is more relevant and robust as compared to SVM-RBF for urban land use classification from VHR imagery in the Indian context. Therefore, finally, Therefore, finally, it is recommended that combining CNN and OBIA is the most promising starting point for further research on developing a robust urban land use classification approach from VHR imagery for developing countries like India.

However, the research question as outlined in this research have been answered below:

- **Specific objective 1: to select suitable image features for urban land use mapping.**

1. What types of image features are extracted from VHR imagery using standard feature extraction methods based on recent literatures?

   In recent literatures, it was observed that GLCM, GLDV, LBP, MPPR and some geometric (e.g., shape, compactness, etc.) and contextual image (e.g., spatial metrics) features were widely used for classifying urban land cover/use from VHR imagery (see section 2 and 3.6.). The previous research has proven that the use of these image features provided good classification results for cities outside of India. Hence, these features have been selected to explore for an Indian city to develop robust features for urban land use classification from VHR imagery (see section 3.6, appendix table C2.1).

2. What is the standard feature selection method used for selecting best features based on recent literatures?

   In recent literatures, several feature selection methods (e.g., PCA, rank, etc.) were explored but SFS with HSIC proven to be a fast and robust feature selection approach in many research (see section 3.11). Hence, SFS-HSIC has been used to select the best features from the large volume of extracted image features.

3. What are best the image features used to map urban land use using standard feature selection method? Using SFS-HSIC, 15 best features (see section 4.7, table 10) such as spectral (image bands, NDVI and mean brightness), textural (GLCM and GLDV) and contextual image features (aggregation index, fractal dimension, cohesion and largest patch index) have been selected from the 135 extracted image features (section 4.5, appendix table C2.1). From this result, it is concluded that the use of handicraft image features for classifying land use from the satellite imagery (e.g., VHR) is not an optimum solution in remote sensing because Hughes phenomena are highly related with high dimensional image features. Hence, feature selection is a robust approach (for a large volume of features) in remote sensing which is prerequisites to develop best classification approach from satellite imagery (e.g., VHR).

- **Specific objective 2: to map urban land uses using SVM and CNN in OBIA and BBIA.**

1. What types of urban land uses are relevant based on national and local land use classification schemes and available literatures?

   In this study, seven relevant urban land use classes such as residential, commercial, industrial, transportation, urban green, vacant land, and waterbody have been selected based on national (NRSA, 2008) and local land use classification schemes (appendix table B5.1 and figure B3.2) and available literatures (section 2, 3.8).

2. What are best parameters of SVM and CNN to improve the classification accuracy of urban land uses employing OBIA and BBIA?

   The best parameters of SVM-RBF (such as gamma 0.0599 and cost is 1000) provided the highest overall accuracy using the best MRS level (level 1 and scale 133) and 150 training samples for OBIA-based urban land use classification. Similarly, the best gamma of 1.292 and cost of 16.681 provided the highest overall accuracy using the best block size (29 by 29 pixel) and 150 training samples for BBIA-based urban land use classification (see section 4.8.1). In addition, best learning and regularization parameters and hyperparameters of CNN provided the highest overall accuracy on 1000 augmented training samples for CNN-BBIA-based urban land use classification (see section 4.8.2). Thus, it is concluded, parameters tuning is one of the best approach in remote sensing to develop robust parameters for classifying land use from satellite imagery (e.g., VHR) because unsuitable parameters always overfitted the classifiers.

3. What are the classification accuracies and time elapses executing a SVM in BBIA and OBIA employing the best parameters and image features?

   The SVM-RBF in OBIA achieved highest overall classification accuracies (e.g., overall accuracy=65.57%, kappa= 0.5467) (see section 4.9.2) as compared to SVM-RBF in BBIA (e.g., overall accuracy=56.64%, kappa= 0.3900) (see section 4.9.3) while employing best features (see section 4.7), parameters and best size of training samples (see section 4.8.1). In addition, the computation time (half an hour in one tile) in both OBIA and BBIA is quite similar. In addition, it is also concluded more time was elapsed for extracting handicraft image features as compared to the classification. Finally, it is concluded that the classification accuracy varied between OBIA and BBIA due to added advantages of addressing geometrical characteristics (shape, size etc.) of land use object in OBIA over the other robust image features and best parameters because urban land use can easily be separated in the VHR imagery in terms of geometry of the land use objects.

4. What are the classification accuracies and time elapses executing a CNN in BBIA employing the best parameters?

   The CNN in BBIA an overall accuracy of 57.51%, and kappa of 0.4245, etc. (see section 4.9.4) employing best parameters (see section 4.8.2). In addition, 5 hours elapse was required to produce a final classified map (one tile) in this classification. From this result, it is concluded that the use of best parameters is the optimum choice to improve the overall classification by reducing the overfitting of the network.

- **Specific objective 3: to evaluate the performance of SVM and CNN in OBIA and BBIA for urban land use classification.**

1. What is the best strategy to measure the accuracy of SVM and CNN for urban land use classification?
   In previous studies, most of the researchers evaluated the performance of classifiers based on only the quantitative agreement, quality of the classified map and time (e.g., Bergado, 2016). However, the best strategy (see section 4.10) is to evaluate the locational agreement and local land use classification scheme along with the quantitative agreement, quality of classified map and time to address the thematic, positional and temporal uncertainty of the classifiers to support local land use classification. Thus, based on the best strategy, other than classical accuracy assessment indices (overall, kappa, etc.), some additional advanced accuracy assessment indices (e.g., recall, precision, F1-score, Klocation, Kno, etc.) were used to evaluate SVM-RBF and CNN (see section 4.10).
2. What is the performance of SVM and CNN for urban land use classification?
   Based on the best strategy adopted for measuring the difference accuracies of SVM-RBF and CNN (see section 4.10), it is concluded that the CNN outperforms the SVM-RBF and similarly OBIA outperforms the BBIA for classifying urban land use from VHR imagery in the Indian context.

5.2. Limitations and contributions

This study has few limitations such as shadow effects in VHR imagery as well as time constraints and limited computer memory storage to experiment on more image tiles as well as across the whole city. This study has several contributions such as firstly, explored the best image features from commonly used huge image features for urban land use mapping. Secondly, explored how to develop best parameters of SVM-RBF and CNN for urban land use classification from VHR imagery. Finally, this study has also explored the best image classification approach (e.g., OBIA) and robust machine learning algorithm (e.g., CNN) for urban land use mapping from VHR imagery for developing countries such as India. This study also provides a starting point for developing detailed land use mapping from VHR imagery to implement better spatial planning policy at the local scale. This is because detailed land use information was hidden due to aggregation pattern of urban land use in the local land use classification scheme. However, in this study, overall classification accuracy somehow low to implement planning policy and in this regard, need some additional research for the improvement of the overall classification accuracy. The additional research has recommended in section 5.3.

5.3. Recommendation for future study

The recommendation of future study has been listed in following heads:

- Use of SVM-RBF for OBIA-based urban land use classification from VHR imagery is highly recommended for Indian context as compared to SVM-RBF with BBIA.
- Use of CNN for BBIA-based urban land use classification from VHR imagery is performing better for the Indian context as compared to SVM-RBF for BBIA.
- Combine CNN with OBIA to develop more a robust urban land use classification approach from VHR imagery for the cities of developing country like India.
- Use of parcels for a BBIA-based urban land use classification is highly recommended as compared to the regular grids to improve the overall accuracy. This will help to extract detail road alignments as well as land use because different attributes of urban land use varies between parcels.
- Use height information such nDSM (e.g., normalized digital surface model) to improve the classification because properties of urban land use varies with the height.
- Experiment on different deep learning network such FCN (e.g., Fully Convolutional Network), DCN (e.g., Deconvolutional Neural Network) and RCNN (e.g., Recurrent Convolutional Neural Network) to learn finer contextual image features to improve the overall classification accuracy and obtain a smoother classification outcome. In addition, experiments on integrating Conditional Random Field (CRF) with CNN is a possible choice to improve the overall classification accuracy (Sun et al., 2016).

LIST OF REFERENCES

Aguilar, M., Vicente, R., Aguilar, F. J., Fernandez, A., & Saldana, M. M. (2012). Optimizing Object-Based Classification in Urban Environments Using Very High Resolution Geoeye-1 Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1–7*(September), 99–104. https://doi.org/10.5194/isprsannals-I-7-99-2012

Ahmed, B., Ahmed, R., & Zhu, X. (2013). Evaluation of Model Validation Techniques in Land Cover Dynamics. *ISPRS International Journal of Geo-Information*, *2*(3), 577–597. https://doi.org/10.3390/ijgi2030577

Alam, S. M. (2011). *Urban growth theories and settlement systems of India*. (K. Markandey & A. Geeta Reddy, Eds.). New Delhi: Concept Pub. Co.

Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction Neural networks in remote sensing. *International Journal of Remote Sensing*, *18*(4), 699–709. https://doi.org/10.1080/014311697218700

Bangalore Development Authority. (2007). Bangalore draft master plan - 2015, An Integrated Planning Approach Towards a Vibrant International City. Retrieved August 15, 2017, from http://www.bdabangalore.org/brochure.pdf

Bergado, J. R. A. (2016). A DEEP FEATURE LEARNING APPROACH TO URBAN SCENE CLASSIFICATION. Enschede, The Netherlands: ITC, Faculty of Geo-information Science and Earth Observation, University of Twente. Retrieved January 10, 2018, from http://www.itc.nl/library/papers_2016/msc/gfm/bergado.pdf

Bergado, J. R. A., Persello, C., & Gevaert, C. (2016). A Deep Learning Approach to the Classification of Sub-Decimetre Resolution Aerial Images. *2016 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 1516–1519. https://doi.org/10.1109/IGARSS.2016.7729387

Berger, C., Voltersen, M., Hese, O., Walde, I., & Schmullius, C. (2013). Robust extraction of urban land cover information from HSR multi-spectral and LiDAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(6), 1–16. https://doi.org/10.1109/JSTARS.2013.2252329

Bruzzone, L., & Persello, C. (2009). Approaches based on support vector machine to classification of remote sensing data. In C. Chen (Ed.), *Handbook of Pattern Recognition and Computer Vision* (pp. 329–352). Singapore: World Scientific.

Camps-valls, G., Mooij, J., & Schölkopf, B. (2010). Kernel Dependence Measures. *Ieee Geoscience and Remote Sensing Letters*, *7*(3), 587–591.

Census. (2011). District Census Handbook. Retrieved February 9, 2018, from http://www.censusindia.gov.in/2011census/dchb/2918_PART_A_DCHB_BANGALORE.pdf

Chadchan, J., & Shankar, R. (2012). An analysis of urban growth trends in the post-economic reforms period in India. *International Journal of Sustainable Built Environment*, *1*(1), 36–49. https://doi.org/10.1016/j.ijsbe.2012.05.001

Chaplin, S. E. (2011). Indian cities, sanitation and the state: the politics of the failure to provide. *Environment and Urbanization*, *23*(1), 57–70. https://doi.org/10.1177/0956247810396277

Chuang, Y.-C., & Shiu, Y.-S. (2016). A Comparative Analysis of Machine Learning with WorldView-2 Pan-Sharpened Imagery for Tea Crop Mapping. *Sensors*, *16*(5), 594. https://doi.org/10.3390/s16050594

Civco, D. L. (1993). Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Systems*, *7*(2), 173–186. https://doi.org/10.1080/02693799308901949

Cockx, K., Van de Voorde, T., & Canters, F. (2014). Quantifying uncertainty in remote sensing-based urban land-use mapping. *International Journal of Applied Earth Observation and Geoinformation*, *31*(1), 154–166. https://doi.org/10.1016/j.jag.2014.03.016

CS231n. (2018). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved February 16, 2018, from http://cs231n.github.io/convolutional-networks/

Dalla Mura, M., Benediktsson, J. A., Waske, B., & Bruzzone, L. (2010). Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing*, *31*(22), 5975–5991. https://doi.org/10.1080/01431161.2010.512425

Damodaran, B. B., Courty, N., & Lefevre, S. (2017). Sparse Hilbert Schmidt Independence Criterion and Surrogate-Kernel-Based Feature Selection for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(4), 2385–2398. https://doi.org/10.1109/TGRS.2016.2642479

Definiens. (2012). Developer XD 2.0.4 Reference Book. Retrieved January 9, 2018, from www.definiens.com

Degeratu, V., Schiopu, P., & Degeratu, S. (2001). McCulloch-Pitts neuron with directional couplers. In *2001 International Semiconductor Conference. CAS 2001 Proceedings (Cat. No.01TH8547)* (Vol. 1, pp. 209–212). IEEE. https://doi.org/10.1109/SMICND.2001.967448

Dias, F., Cousty, J., & Najman, L. (2014). Dimensional operators for mathematical morphology on simplicial complexes. *Pattern Recognition Letters*, *47*, 111–119. https://doi.org/10.1016/J.PATREC.2014.01.006

Drăguţ, L., Csillik, O., Eisank, C., & Tiede D. (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing*, *88*, 119–127. https://doi.org/10.1016/j.isprsjprs.2013.11.018

Drăguţ, L., & Eisank, C. (2012). Automated object-based classification of topography from SRTM data. *Geomorphology*, *141–142*, 21–33. https://doi.org/10.1016/j.geomorph.2011.12.001

Drăguţ, L., Tiede, D., & Levick, S. R. (2010). ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, *24*(6), 859–871. https://doi.org/10.1080/13658810903174803

Duque, J. C., Patino, J. E., & Betancourt, A. (2017). Exploring the potential of machine learning for automatic slum identification from VHR imagery. *Remote Sensing*, *9*(9), 1–23. https://doi.org/10.3390/rs9090895

Gevaert, C. M., Persello, C., Sliuzas, R., & Vosselman, G. (2017). Informal settlement classification using point-cloud and image-based features from UAV data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *125*, 225–236. https://doi.org/10.1016/j.isprsjprs.2017.01.017

Gevaert, C. M., Persello, C., & Vosselman, G. (2016). Optimizing multiple kernel learning for the classification of UAV data. *Remote Sensing*, *8*(12). https://doi.org/10.3390/rs8121025

Gokon, H., Post, J., Stein, E., Martinis, S., Twele, A., Muck, M., … Matsuoka, M. (2015). A Method for Detecting Buildings Destroyed by the 2011 Tohoku Earthquake and Tsunami Using Multitemporal TerraSAR-X Data. *IEEE Geoscience and Remote Sensing Letters*, *12*(6), 1277–1281. https://doi.org/10.1109/LGRS.2015.2392792

González-Castro, V., Debayle, J., & Pinoli, J.-C. (2014). Color Adaptive Neighborhood Mathematical Morphology and its application to pixel-level classification. *Pattern Recognition Letters*, *47*, 50–62. https://doi.org/10.1016/J.PATREC.2014.01.007

Goodfellow, I., and Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press. Retrieved 24 February, 2018, from http://www.deeplearningbook.org

Haralick, R., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. https://doi.org/10.1109/TSMC.1973.4309314

Herold, M., Couclelis, H., & Clarke, K. C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*, *29*(4), 369–399. https://doi.org/10.1016/j.compenvurbsys.2003.12.001

Herold, M., Liu, X., & Clark, K. C. (2003). Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering and Remote Sensing*, *69*(9), 991–1001. https://doi.org/10.14358/PERS.69.9.991

Hu, F., Xia, G.-S., Hu, J., & Zhang, L. (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing*, *7*(11), 14680–14707. https://doi.org/10.3390/rs71114680

Hu, S., & Wang, L. (2013). Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing*, *34*(3), 790–803. https://doi.org/10.1080/01431161.2012.714510

Huang, X., & Zhang, L. (2012). Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(1), 161–172. https://doi.org/10.1109/JSTARS.2011.2168195

Huang, Y., Zhao, C., Yang, H., Song, X., Chen, J., & Li, Z. (2017). Feature selection solution with high dimensionality and low-sample size for land cover classification in object-based image analysis. *Remote Sensing*, *9*(9), 1–16. https://doi.org/10.3390/rs9090939

Humboldt State University. (2018). GSP2016-Introduction to Remote Sensing. Retrieved March 5, 2018, from http://gsp.humboldt.edu/olm_2016/Courses/GSP_216_Online/lesson6-2/metrics.html

Kamal, M., Phinn, S., & Johansen, K. (2015). Object-Based Approach for Multi-Scale Mangrove Composition Mapping Using Multi-Resolution Image Datasets. *Remote Sensing*, *7*(4), 4753–4783. https://doi.org/10.3390/rs70404753

Kim, E., Lee, S., Kim, J., Kim, C., Byun, Y., Kim, H., & Lee, T. (2012). Pattern Recognition for Selective Odor Detection with Gas Sensor Arrays. *Sensors*, *12*(12), 16262–16273. https://doi.org/10.3390/s121216262

Kim, M.-C., & Choi, C.-H. (1997). A New Weight Initialization Method for the MLP with the BP in Multiclass Classification Problems. *Neural Processing Letters*, *6*, 11–23.

Kim, T. (2010). Pattern Recognition Using Artificial Neural Network: A Review. In Y. X. Samir Kumar Bandyopadhyay, Wael Adi, Tai-hoon Kim (Ed.), *Information Security and Assurance* (pp. 138–148). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13365-7_14

Kohli, D. (2015). *Identifying and classifying slum areas using remote sensing (Doctoral dissertation).* https://doi.org/10.3990/1.9789036540087

Kotharkar, R., Bahadure, P., & Sarda, N. (2014). Measuring compact urban form: A case of Nagpur city, India. *Sustainability (Switzerland)*, *6*(7), 4246–4272. https://doi.org/10.3390/su6074246

Kuffer, M., & Barros, J. (2011). Urban morphology of unplanned settlements: The use of spatial metrics in VHR remotely sensed images. *Procedia Environmental Sciences*, *7*, 152–157. https://doi.org/10.1016/j.proenv.2011.07.027

Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I., & Maarseveen, M. (2017). Capturing the Diversity of Deprived Areas with Image-Based Features: The Case of Mumbai. *Remote Sensing*, *9*(4), 384. https://doi.org/10.3390/rs9040384

Laliberte, A. S., & Rango, A. (2009). Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 1–10. https://doi.org/10.1109/TGRS.2008.2009355

Lee, H., & Kwon, H. (2016). Going Deeper with Contextual CNN for Hyperspectral Image Classification, *26*(10), 1–14. https://doi.org/10.1109/TIP.2017.2725580

Li, M. (2016). *Urban Land Use Extraction From Very High Resolution Remote Sensing Images By Bayesian Network*. Retrieved February 15, 2018, from https://www.itc.nl/library/papers_2017/phd/mmli.pdf

Li, M., Stein, A., Bijker, W., & Zhan, Q. (2016). Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 192–205. https://doi.org/10.1016/j.isprsjprs.2016.10.007

Liao, W., Chanussot, J., Dalla Mura, M., Huang, X., Bellens, R., Gautama, S., & Philips, W. (2017). Taking optimal advantage of fine spatial resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images. *IEEE Geoscience and Remote Sensing Magazine*, *5*(2), 8–28. https://doi.org/10.1109/MGRS.2017.2663666

Liao, W., Mura, M. D., Chanussot, J., Bellens, R., Philips, W., & Member, S. (2015). With Partial Reconstruction, 1–19. https://doi.org/10.1109/TGRS.2015.2488280

Lin, Y.-J. (2011). Prevention of transient instability employing rules based on back propagation based ANN for series compensation. *International Journal of Electrical Power & Energy Systems*, *33*(10), 1776–1783. https://doi.org/10.1016/J.IJEPES.2011.09.001

Liu, H., Huang, X., Wen, D., & Li, J. (2017). The use of landscape metrics and transfer learning to explore urban villages in China. *Remote Sensing*, *9*(4), 1–23. https://doi.org/10.3390/rs9040365

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*(5), 823–870. https://doi.org/10.1080/01431160600746456

Ma, L., Cheng, L., Li, M., Liu, Y., & Ma, X. (2015). Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *102*, 14–27. https://doi.org/10.1016/j.isprsjprs.2014.12.026

Ma, L., Ma, F., Ji, Z., Gu, Q., Wu, D., Deng, J., & Ding, J. (2015). Urban Land Use Classification Using LiDAR Geometric, Spatial Autocorrelation and Lacunarity Features Combined with Postclassification Processing Method. *Canadian Journal of Remote Sensing*, *41*(4), 334–345. https://doi.org/10.1080/07038992.2015.1102630

Madani, K. (2008). Artificial Neural Networks Based Image Processing & Pattern Recognition: From Concepts to Real-World Applications. In *2008 First Workshops on Image Processing Theory, Tools and Applications* (pp. 1–9). IEEE. https://doi.org/10.1109/IPTA.2008.4743797

Man, Q., Dong, P., & Guo, H. (2015). Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification. *International Journal of Remote Sensing*, *36*(March), 1618–1644. https://doi.org/10.1080/01431161.2015.1015657

MatConvNet. (2018). MatConvNet: CNNs for MATLAB. Retrieved March 2, 2018, from http://www.vlfeat.org/matconvnet/

Mboga, N. O. (2017). *Detection of Informal Settlements From Vhr Satellite Images Using Convolutional Neural Networks*. Retrieved March 15, 2018, from https://webapps.itc.utwente.nl/librarywww/papers_2017/msc/gfm/mboga.pdf

McGarigal, K., Cushman, SA., & Ene, E. (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Retrieved January 9, 2018, from http://www.umass.edu/landeco/research/fragstats/fragstats.html

Megahed, Y., Cabral, P., Silva, J., & Caetano, M. (2015). Land Cover Mapping Analysis and Urban Growth Modelling Using Remote Sensing Techniques in Greater Cairo Region—Egypt. *ISPRS International Journal of Geo-Information*, *4*(3), 1750–1769. https://doi.org/10.3390/ijgi4031750

Mourão-Miranda, J., Hardoon, D. R., Hahn, T., Marquand, A. F., Williams, S. C. R., Shawe-Taylor, J., & Brammer, M. (2011). Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine. *NeuroImage*, *58*(3), 793–804. https://doi.org/10.1016/j.neuroimage.2011.06.042

Naorem, V., Kuffer, M., Verplanke, J., & Kohli, D. (2016). Robustness of rule sets using VHR imagery to detect informal settlements - a case of Mumbai, India. *GEOBIA 2016 : Solutions and Synergies*. https://doi.org/10.3990/2.416

Niazmardi, S., Safari, A., & Homayouni, S. (2017). Similarity-Based Multiple Kernel Learning Algorithms for Classification of Remotely Sensed Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 2012–2021. https://doi.org/10.1109/JSTARS.2017.2662484

Nijman, J. (2008). Against the odds: Slum rehabilitation in neoliberal Mumbai. *Cities*, *25*(2), 73–85. https://doi.org/10.1016/J.CITIES.2008.01.003

Nikolakopoulos, K., & Oikonomidis, D. (2015). Quality assessment of ten fusion techniques applied on worldview-2. *European Journal of Remote Sensing*, *48*(September), 141–167. https://doi.org/10.5721/EuJRS20154809

Nikolakopoulos, K., & Konstantinos, G. (2008). Comparison of Nine Fusion Techniques for Very High Resolution Data. *Photogrammetric Engineering & Remote Sensing*, *74*(5), 647–659.

Nikolay S. (2017). Local binary patterns - File Exchange - MATLAB Central. Retrieved January 14, 2018, from https://nl.mathworks.com/matlabcentral/fileexchange/36484-local-binary-patterns

Nouri, H., Beecham, S., Anderson, S., & Nagler, P. (2013). High spatial resolution WorldView-2 imagery for mapping NDVI and its relationship to temporal urban landscape evapotranspiration factors. *Remote Sensing*, *6*(1), 580–602. https://doi.org/10.3390/rs6010580

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623

Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2017). A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2017.11.021

Park, M. H., & Stenstrom, M. K. (2008). Classifying environmentally significant urban land uses with satellite imagery. *Journal of Environmental Management*, *86*(1), 181–192. https://doi.org/10.1016/j.jenvman.2006.12.010

Pathan, S. K., Jothimahi, P., Kumar, D. S., & Pendharkar, S. P. (1989). Urban land use mapping and zoning of Bombay Metropolitan Region using remote sensing data. *Journal of the Indian Society of Remote Sensing*, *17*(3), 11–22. https://doi.org/10.1007/BF02995827

Pathan, S., Shukla, V., Patel, R., Patel, B., & Mehta, K. (1991). Urban land use mapping: A case study of ahmedabad city and its environs. *Journal of the Indian Society of Remote Sensing*, *19*(2), 95–112. https://doi.org/10.1007/BF03008125

Pedersen, G. B. M. (2016). Semi-automatic classification of glaciovolcanic landforms: An object-based mapping approach based on geomorphometry. *Journal of Volcanology and Geothermal Research*, *311*, 29–40. https://doi.org/10.1016/J.JVOLGEORES.2015.12.015

Persello, C., & Bruzzone, L. (2010a). A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(3 PART 1), 1232–1244. https://doi.org/10.1109/TGRS.2009.2029570

Persello, C., & Bruzzone, L. (2014b). Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(11), 6937–6956. https://doi.org/81800c 10.1117/12.898483

Persello, C., & Bruzzone, L. (2016c). Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2615–2626. https://doi.org/10.1109/TGRS.2015.2503885

Pietikäinen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*. (M. VIERGEVER, Ed.) (Vol. 40). Springer-Verlag London Limited. https://doi.org/10.1007/978-0-85729-748-8

Pratomo, J. (2016). Transferability of The Generic and Local Ontology of Slum in Multi-temporal Imagery, Case Study: Jakarta, (March), 90. Retrieved August 12, 2017, from http://www.itc.nl/library/papers_2016/msc/upm/pratomo.pdf

Radoux, J., & Bogaert, P. (2017). Good practices for object-based accuracy assessment. *Remote Sensing*, *9*(7). https://doi.org/10.3390/rs9070646

Richards, J. a, & Jia, X. (2006). *Remote Sensing Digital Image Analysis*. *Methods*. https://doi.org/10.1007/3-540-29711-1

Rotated Local Binary Pattern (RLBP): Rotation invariant texture descriptor. (n.d.). Retrieved January 12, 2018, from http://www.cs.tut.fi/~mehta/rlbp

Sameen, M. I., & Pradhan, B. (2016). A Simplified Semi-Automatic Technique for Highway Extraction from High-Resolution Airborne LiDAR Data and Orthophotos. *Journal of the Indian Society of Remote Sensing*, *45*(3), 1–11. https://doi.org/10.1007/s12524-016-0610-5

Sandborn, A., & Engstrom, R. N. (2016). Determining the Relationship between Census Data and Spatial Features Derived from High-Resolution Imagery in Accra, Ghana. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1970–1977. https://doi.org/10.1109/JSTARS.2016.2519843

Shabat, A. M., & Tapamo, J. R. (2014). A comparative study of Local directional pattern for texture classification. In *2014 World Symposium on Computer Applications & Research (WSCAR)* (pp. 1–7). IEEE. https://doi.org/10.1109/WSCAR.2014.6916773

Silván-Cárdenas, J. L., Almazán-González, J. A., & Couturier, S. A. (2014). Remote Identification of Housing Buildings with High-Resolution Remote Sensing (pp. 380–390). NewYork: Springer, Cham. https://doi.org/10.1007/978-3-319-07491-7_39

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Stavrakoudis, D. G., Dragozi, E., Gitas, I. Z., & Karydas, C. G. (2014). Decision fusion based on hyperspectral and multispectral satellite imagery for accurate forest species mapping. *Remote Sensing*, *6*(8), 6897–6928. https://doi.org/10.3390/rs6086897

Stow, D., Lopez, A., Lippitt, C., Hinton, S., & Weeks, J. (2007). Object-based classification of residential land use within Accra, Ghana based on QuickBird satellite data. *International Journal of Remote Sensing*, *28*(22), 5167–5173. https://doi.org/10.1080/01431160701604703

Sun, X., Lin, X., Shen, S., & Hu, Z. (2016). High-Resolution Remote Sensing Data Classification over Urban Areas Using Random Forest Ensemble and Fully Connected Conditional Random Field, 1–22. https://doi.org/10.3390/ijgi6080245

Suykens, J. A. K. (2001). Support Vector Machines: A Nonlinear Modelling and Control Perspective. *European Journal of Control*, *7*(2–3), 311–327. https://doi.org/10.3166/ejc.7.311-327

Tang, Q., Wang, L., Li, B., & Yu, J. (2012). Towards a comprehensive evaluation of V-I-S sub-pixel fractions and land surface temperature for urban land-use classification in the USA. *International Journal of Remote Sensing*, *33*(19), 5996–6019. https://doi.org/10.1080/01431161.2012.675453

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, *160*, 1–14. https://doi.org/10.1016/j.rse.2015.01.006

United Nations. (2015). World Urbanization Prospects. *United Nations*, *12*, 32. https://doi.org/10.4054/DemRes.2005.12.9

Wang, Y., Zhao, Y., & Chen, Y. (2014). Texture classification using rotation invariant models on integrated local binary pattern and Zernike moments. *EURASIP Journal on Advances in Signal Processing*, *2014*(1), 182. https://doi.org/10.1186/1687-6180-2014-182

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Wentz, E. ., Nelson, D. ., Rahman, A. ., Stefanov, W. ., & Roy, S. S. . (2008). Expert system classification of urban land use/cover for Delhi, India. *International Journal of Remote Sensing*, *29*(15), 4405–4427. https://doi.org/10.1080/01431160801905497

Wieland, M., & Pittore, M. (2014). Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images. *Remote Sensing*, *6*(4), 2912–2939. https://doi.org/10.3390/rs6042912

Wu, B., Xiong, Z. guo, Chen, Y. zhi, & Zhao, Y. di. (2009). Classification of quickbird image with maximal mutual information feature selection and support vector machine. *Procedia Earth and Planetary Science*, *1*(1), 1165–1172. https://doi.org/10.1016/j.proeps.2009.09.179

Wu, S., Silvánhyphen;Cárdenas, J., & Wang, L. (2007). Per‐field urban land use classification based on tax parcel boundaries. *International Journal of Remote Sensing*, *28*(12), 2777–2801. https://doi.org/10.1080/01431160600981541

Yanchen, W., Yinghai, K., Huili, G., Beibei, C., & Lin, Z. (2014). Comparison of object-based and pixel-based methods for urban land-use classification from WorldView-2 imagery. *3rd International Workshop on Earth Observation and Remote Sensing Applications, EORSA 2014 - Proceedings*, *2*, 284–288. https://doi.org/10.1109/EORSA.2014.6927896

Yang, H., Ma, B., Du, Q., & Yang, C. (2010). Improving urban land use and land cover classification from high-spatial-resolution hyperspectral imagery using contextual information. *Journal of Applied Remote Sensing*, *4*(1), 41890. https://doi.org/10.1117/1.3491192

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 Sensing*, *1*(3), 243–265. https://doi.org/10.3390/rs1030243

Yusuf, Y., Sri Sumantyo, J. T., & Kuze, H. (2013). Spectral information analysis of image fusion data for remote sensing applications. *Geocarto International*, *28*(4), 291–310. https://doi.org/10.1080/10106049.2012.692396

Zhang, L., & Zhang, B. (1999). A geometrical representation of McCulloch-Pitts neural model and its applications. *IEEE Transactions on Neural Networks*, *10*(4), 925–929. https://doi.org/10.1109/72.774263

Zhan, Q., Molenaar, M., & Xiao, Y. (2001). Hierarchical object-based image analysis of high-resolution imagery for urban land use classification. *Remote Sensing and Data Fusion over Urban Areas, IEEE/ISPRS Joint Workshop 2001*, 35–39. https://doi.org/10.1109/DFUA.2001.985721

Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., & Wang, W. (2017). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment and Urban Systems*, *64*, 215–228. https://doi.org/10.1016/j.compenvurbsys.2017.03.001

Zhang, X., Cui, J., Wang, W., & Lin, C. (2017). A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors (Switzerland)*, *17*(7). https://doi.org/10.3390/s17071474

Zhen, Z., Quackenbush, L. J., Stehman, S. V., & Zhang, L. (2013). Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *International Journal of Remote Sensing*, *34*(19), 6914–6930. https://doi.org/10.1080/01431161.2013.810822

APPENDIX

A. Review of literature

A1. Hyperparameters of supervised CNN network

Table A1.1. Hyperparameters of supervised CNN network (Bergado, 2016; Goodfellow et al., 2016; Mboga, 2017).

Hyperparameters	Descriptions
(A) Learning and regularization parameters	
i. Batch size	Number of training samples train by CNN in each iteration.
ii. Maximum number of epoch, τ	Number of iteration during training the CNN network.
iii. Learning rate, ϵ	Learning fraction which explained how precisely updates the derivates of the SGD to obtain minimum loss function in each epoch.
iv. Momentum, α	Fraction used to accelerate the network during training.
v. Weight decay, λ	It is used in L2 regularization norm which performs new loss function to minimize the overfitting of network.
vi. Dropout, d_r	It is one of the very important regularization parameter which explained as the percent of randomly dropping (d_r) of neuron which is co-adapted with the other neuron. The co-adaptation of neurons leads to the overfitting of the network.
(B) CNN configuration parameters	
vii. Patch size	Square patch which is 2-D image grid learn by network from the image.
viii. Number of patch	Number of 2-D image grid learn by network from the image.
ix. Filter size	Dimension of 2-D filters in convolutional layers used to develop activation features from the 2-D image patches learned by network.
x. Number of filter	Number of 2-D filters in convolutional layers used to develop activation features from the learned 2-D image patches.
xi. Pooling size	Downsampling of non-linear activation feature in convolutional layers as a function of max pooling which extract the maximum value in the n-size of pooling region.
xii. Network depth	Number of convolutional layers and FC layers are termed as the depth of the CNN network which affects the overall classification accuracy.

A2. Multiresolution image segmentation

Table A2.1. Literatures review on selecting scale parameter, shape and compactness

Sl. No.	Scale parameters	Compactness	Shape	Applications	Overall Accuracy (%)	References
1	Default ESP2	0.50	0.50	Informal settlement mapping	85.54-86.33	(Naorem etal., 2016)
2	ESP2 (55-153)	0.50	0.50	Slum mapping	64.00-70.80	(Pratomo, 2016)
3	ESP2 (40-300)	0.50	0.50	Slum mapping	47.00-68.00	(Kohli, 2015)
4	50	0.50	0.50	Urban land use mapping	79.60-96.0	(Yanchen et al., 2014)
5	10-150	0.50	0.70	Residential land use mapping	75.00	(Stow et al., 2007)

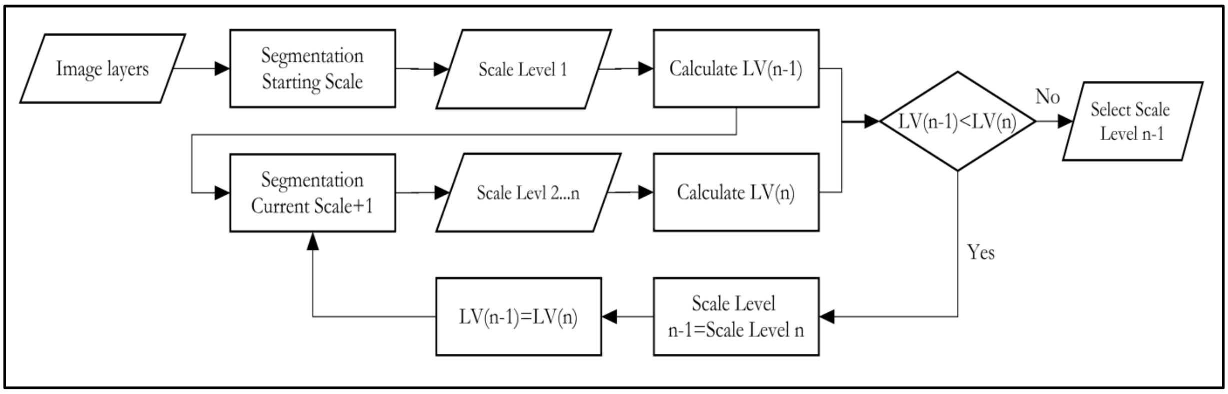


Figure A2.1. Process of ESP tool for the estimation of scale parameter based on threshold of local variance (Drăgut et al., 2014)

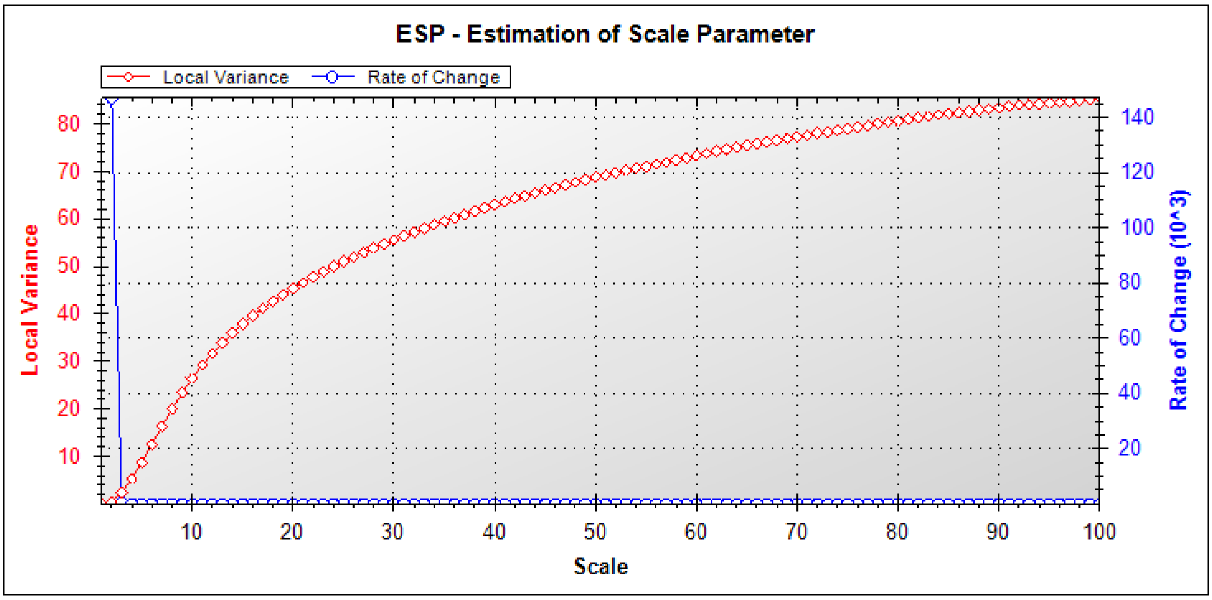


Figure A2.2. LV and ROC estimated using ESP tool for three-level hierarchy MRS of Worldview fused imagery, 2015

B1. Methodological flowchart

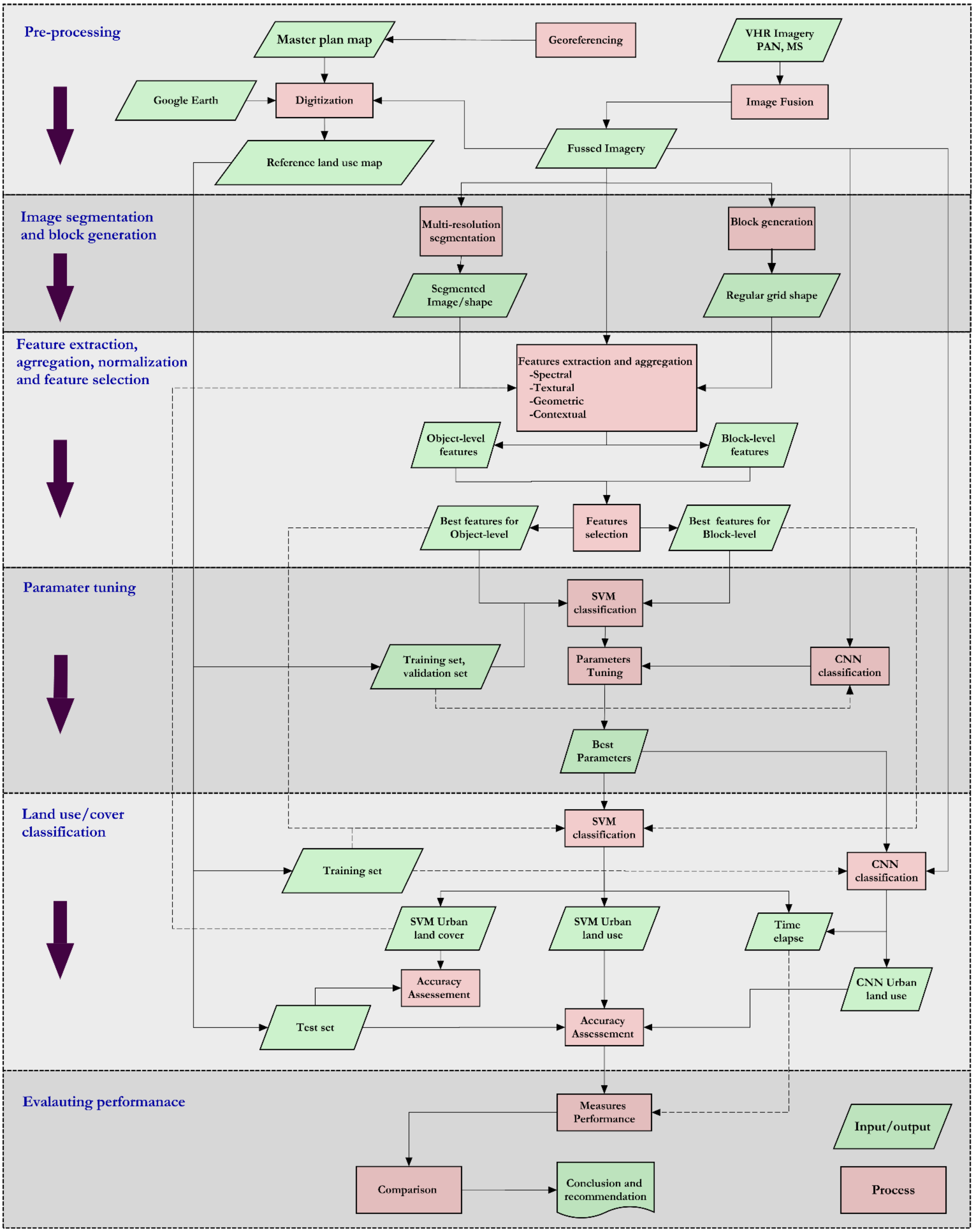


Figure B1.1. Shows methodological flowchart

B2. Study area

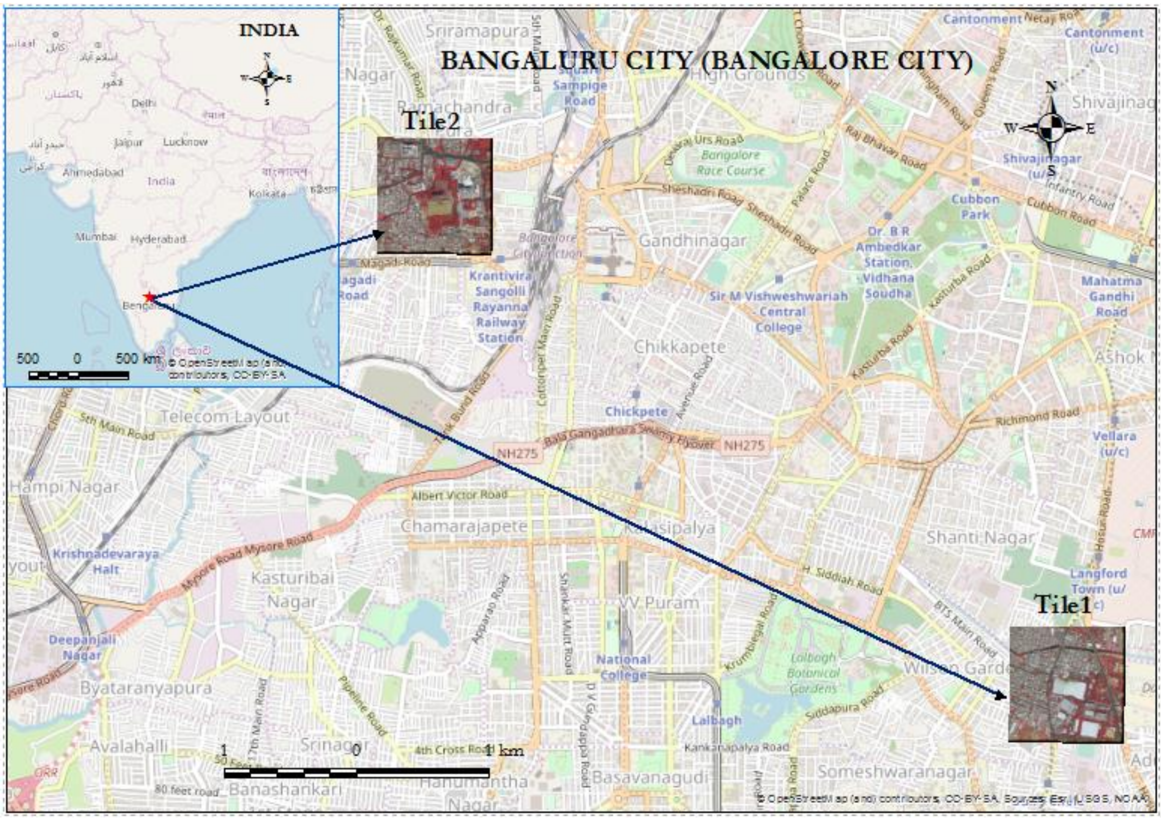


Figure B2.1. Shows location of the study area (source: Openstreet map)

B3. Dataset and software used

Table B3.1. Worldview3 Satellite imagery

Sensors	Spectral Bands (µm)	Spatial resolution (metre)	Radiometric resolution	Off nadir	Sun elevation	Repeat cycle	Swath (km)
PAN	Band PAN: 0.45-0.80	0.34	11 bits	22.79^0	52.79^0	1 day	13.10
MS	Band coastal: 0.40-0.45	1.38	14 bits				
	Band blue: 0.45-0.51						
	Band green: 0.51-0.58						
	Band yellow: 0.585-0.625						
	Band red: 0.63-0.69						
	Band red edge:0.705-0.745						
	Band NIR1: 0.77-0.895						
	Band NIR2: 0.86-1.02						

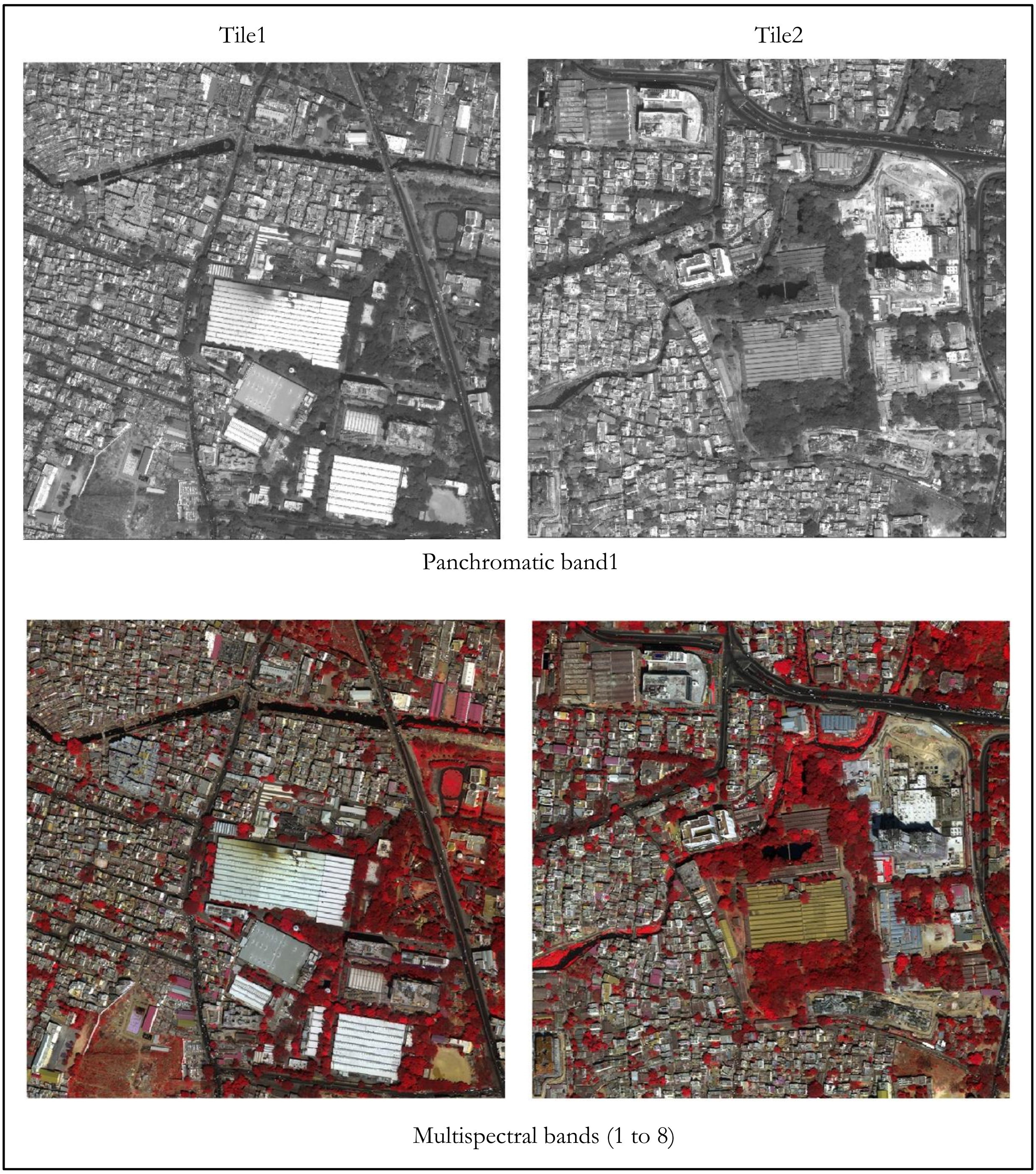


Figure B3.1. Shows original Worldview3 satellite imageries of tile1 and tile2 with 2501 X 2501 pixels dimension, 2015

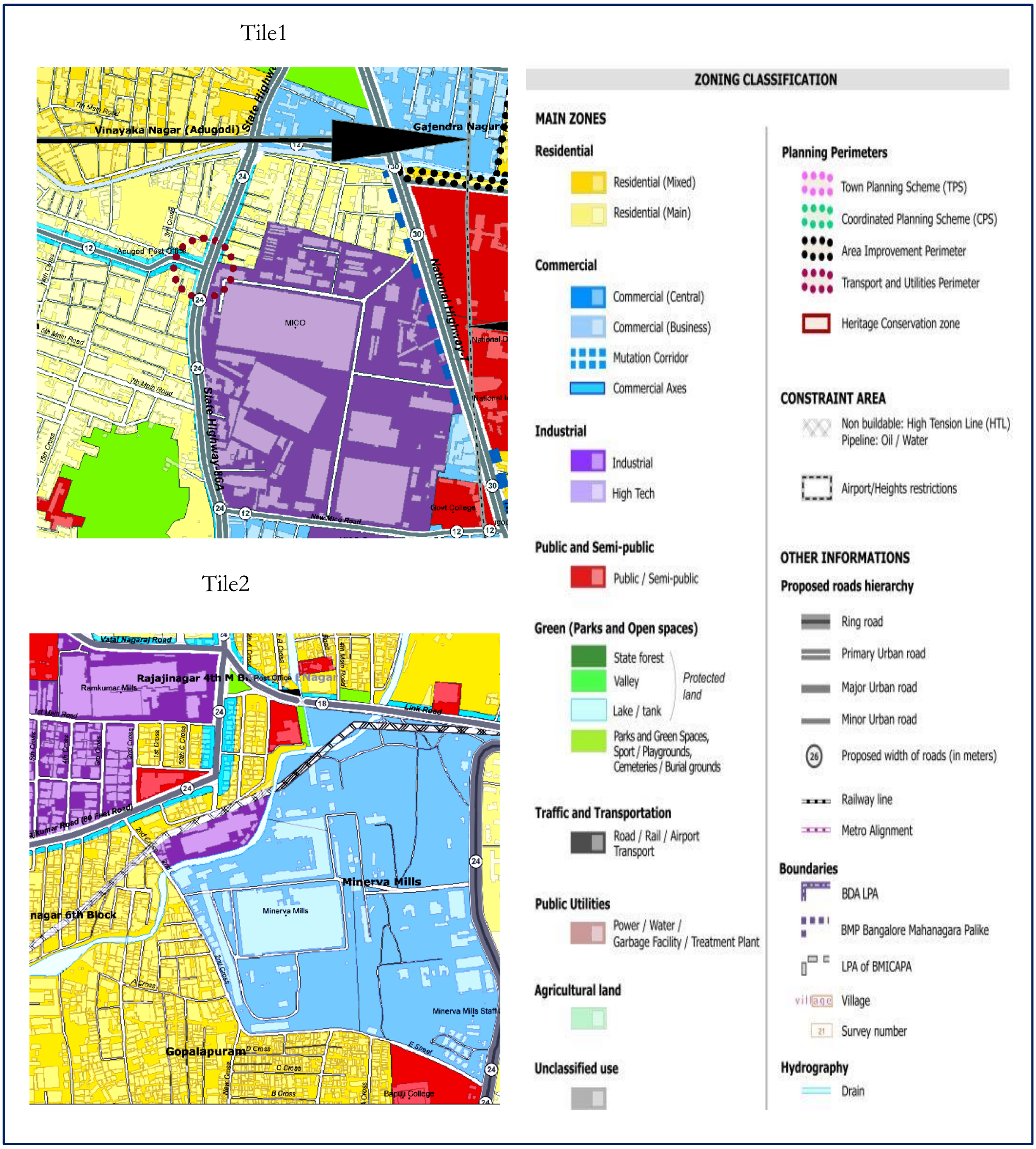


Figure B3.2. Shows master plan map of tile1 and tile2 with 2501 X 2501 pixels dimension, 2015

Table B3.2. Hardware and Software used for this study

Sl.no	Hardware and operating system	software	Purpose of use
1	• Hp-Pavilion desktop 7010TX, CPU-intel corei7, DDR3, RAM-12GB, HDD-640 GB, Graphic-NVidia 2GB, OS-window7 home premium and • DELL OptiPlex &Precision Workstation CPU- Xeon E5-2643, Core 6, RAM-128GB, HDD- 2 TB, SSD-1TB, OS- Linux Ubuntu	Arc GIS 10.5.1	• Creation of fishnet regular grid as the block, • Labelled block and objects for training and test sets, • Creates features in raster format from the features extracted in shape file and • Layout map design
2		Erdas Imagine 2015	• HPF image fusion, • Subset image, geo-referencing and • Feature aggregation with zonal statistics
3		ECognition 9.2	• Multi-resolution and chessboard segmentation and • Feature extraction and aggregation
		Fragstats 4.2.1	• Developed contextual features based on spatial matrices
4		R Studio 3.4.1	• SFS feature section, • SVM-RBF parameter tuning for urban land cover and urban land use classification and • SVM-RBF urban land cover and urban land use classification
5		Matlab 2017a&2017b	• Extraction of LBF and MPPR features, • Export of file from Matlab format to ENVI format • CNN development • CNN for Urban land use classification
6		ENVI classic 5.3	• Image normalization
7		Idrisi Selva 17.02	• Measuring location-based accuracy indices
8		Microsoft office 2016	• Thesis writing, chart and diagram, tabulation
9		Adobe acrobat 8	• Conversion of word file to pdf
10		NCH software	• Flowchart for methodology

B4. Geometric features

Table B4.1. Description of geometric features used for OBIA-based Urban land cover/use classification (Definiens, 2012)

No.	Geometric features	Descriptions
1	Asymmetry	Object is separated based on the comparison of the asymmetrical shape which describes the relative length of the object compared to regular polygon, measured by the ratio of major and minor axis of ellipse (e.g., square and rectangular building, road, urban green polygon, etc.).
2	Compactness	Object is separated based on the compactness of the object.
3	Elliptical fit	Object is separated by comparing shape of the object with the elliptical shape (e.g., square building from urban green polygon etc.).
4	Rectangular fit	Object is separated by comparing the fitness of the object with the rectangular shape (e.g., square building from rectangular building etc.).
5	Roundness	Object is separated by comparing how similar the object is to an ellipse which is calculated based by the radius of enclosed and enclosing ellipse (e.g., different polygons of waterbody, urban green, vacant land etc.).
6	Main direction	Object is separated by comparing directional change which is measured by two larger eigenvalues of spatial distribution of object (e.g., road from river etc.).
7	Shape Index	Object is separated based on the different shapes (e.g., square and rectangular building, road, urban green, etc.).
8	Border Index	Object is separated based on the jaggedness of the object which is calculated by using rectangular approximation (e.g., square building from rectangular building, road etc.).
9	Radius of largest enclosed ellipse	Object is separated by comparing how object is similar to an ellipse (e.g., different polygons of waterbody, urban green, vacant land etc.).
10	Radius of smallest enclosing ellipse	Object is separated by comparing how much the shape of the object is similar to an ellipse (e.g., different polygons of waterbody, urban green, vacant land etc.).
11	Border length	Object is separated based on the total length of the edges of the object shared with the edge of the other object (e.g., building, road, urban green etc.).
12	Width	Object is separated based on the width which is calculated from length to width ratio of the object (e.g., road from the building).
13	Length-width ratio	Object is separated based on the length and width ratio (e.g., road from the building or square building from rectangular building).
14	Number of pixels	Number of pixels help to separate small object from the big object (e.g., small buildings from large buildings).

B5. Proposed urban land cover/use classification scheme

Table B5.1. Proposed urban land cover/land use classification scheme

Class id	Urban Land cover	Urban Land use		
		Class id	Types	Descriptions
1	Built-up	1	Residential	This includes all types of formal and informal settlements used for living. Mixed and public and semi-public land uses are included in the residential land use because of their similar building and infrastructural characteristics.
		2	Commercial	Used for business, trade and commercial activity. Large commercial area located in the city centre and close to highly populated area. It has large building size, and complex shape.
		3	Industrial	Used for production, manufacturing, factory, warehousing etc. which is located close to rail line, and commercial area.
2	Road	4	Transportation	Used for transportation. It is linear shape, which includes rail, road, railway station and bus stops.
3	Vegetation	5	Urban green	It includes parks, trees, plantation and agriculture.
4	Undeveloped	6	Vacant land	It includes undeveloped land, area under construction and cemetery etc.
5	Waterbody	7	Waterbody/ shadow	Canal, ponds tanks, river etc.
6	Shadow			Shades of different image objects (buildings/tress).

B6. Feature selection

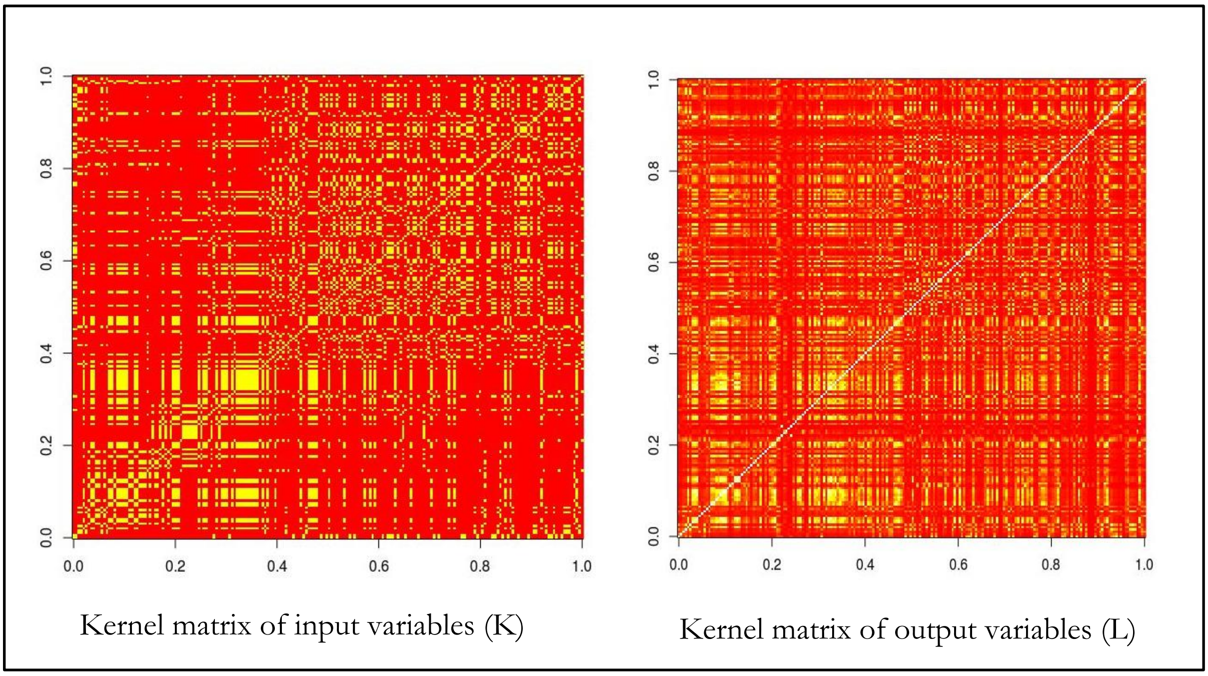


Figure B6.1. Shows kernel matrices of input variables (K) and output variables (L) associated with HSIC=0.058 and best sigma=0.464 in this study. The best sigma is selected based on the maximum HSIC value which was evaluated using 10 different sequential sigma values that varies from 0.10 and 100 (using sigma<-10^seq(-1,2, len=10)).

C. Results and Discussions

C1. Referenced land cover maps

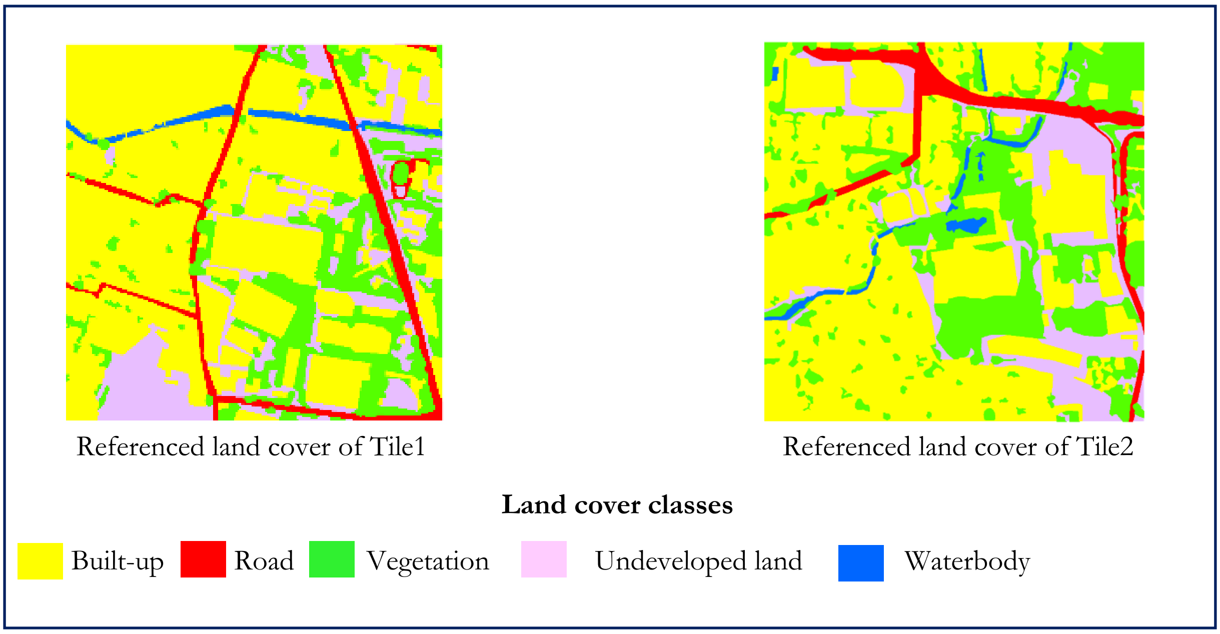


Figure C1.1. Show referenced urban land cover maps, 2015

C2. Extracted, aggregated and normalized image features

Table C2.1. Extracted image features

Sl.no	Image features	Methods	Image features	Number of features	Sources
1	Spectral	Mean	Mean of all spectral bands	08	Fused VHR image
			Mean brightness	01	
		NDVI	NDVI feature	01	
2	Textural	GLCM	GLCM (mean)	08	
			GLCM (variance)	08	
			GLCM (homogeneity)	08	
			GLCM (contrast)	08	
			GLCM (dissimilarity)	08	
			GLCM (entropy)	08	
			GLCM (second moment)	08	
			GLCM (correlation)	08	
		GLDV	GLDV (ang.2 moment)	08	
			GLDV(entropy)	08	
			GLDV(mean)	08	
			GLDV(contrast)	08	
		LBP	LBP feature	08	
		MPPR	MPPR feature	02	Panchromatic
3	Geometric	Object level geometry	Asymmetry	01	Fused VHR image
			Compactness	01	
			Elliptical fit	01	
			Rectangular fit	01	
			Roundness	01	
			Main direction	01	
			Shape Index	01	
			Boarder Index	01	
			Radius of largest enclosed ellipse	01	
			Radius of smallest enclosed ellipse	01	
			Border length	01	
			Width	01	
			Length-width ratio	01	
			Number of pixels	01	
4	Contextual	Spatial metrics	Path density	01	Urban Land cover (built-up)
			Aggregation index	01	
			Fractal dimension	01	
			Cohesion	01	
			Largest patch Index	01	
Total features				135	-

C3. Training and test samples

C3.1. Training and test samples for SVM-RBF-OBIA-based urban land cover classification

Table C3.1.1. Training and test samples at MRS level

MRS level	Total segments	Total training samples	Splitting training samples into 60:40		Test samples whole tile
			Training set	Validation set	
Level 1	3171	200	120	80	6255001
Level 2	1100				
Level 3	508				

Table C3.1.2. Training and test samples for SVM-RBF-OBIA-based urban land cover classification (MRS level 1)

Class ID	Training land cover classes	Tile 1		Tile 2	
		Number of training objects	Number of test pixel	Number of training objects	Number of test pixel
1	Building	75	3659387	75	3307488
2	Road	20	369708	20	344350
3	Vegetation	46	1129351	46	1625177
4	Undeveloped land	23	895471	23	826730
5	Water body	10	99406	10	85615
6	Shadow	26	101678	26	65641
Total		200	6255001	200	6255001

C3.2. Training and test samples for SVM-RBF-OBIA-based urban land use classification

Table C3.2.1. Training and test samples at MRS level

MRS level	Total segments	Total training samples	Splitting training samples into 60:40		Test pixels whole tile
			Training set	Validation set	
Level 1	3171	200	120	80	6255001
Level 2	1100				
Level 3	508				

Table C3.2.2. Training and test samples for SVM-RBF-OBIA-based urban land use classification (MRS level 2)

Class ID	Training land use classes	Tile1	
		Number of training objects	Number of test pixels
1	Residential	58	2717828
2	Commercial	15	184944
3	Industrial	40	845305
4	Transportation	16	359030
5	Urban green	35	1099990
6	Vacant land	13	873329
7	Water body/Shadow	23	174575
Total		200	6255001

Table C3.2.3. Training and test samples for SVM-RBF-OBIA-based urban land use classification (MRS level 3)

Class ID	Training land use classes	Tile1	
		Number of training objects	Number of test pixels
1	Residential	66	2717828
2	Commercial	14	184944
3	Industrial	50	845305
4	Transportation	12	359030
5	Urban green	26	1099990
6	Vacant land	13	873329
7	Water body/Shadow	19	174575
Total		200	6255001

C3.3. Training and test samples for SVM-RBF-BBIA-based urban land use classification

Table C3.3.1. Training and test samples at block level

Blocks size	Total blocks	Total training samples	Splitting training samples into 60:40		Test pixels whole tile
			Training set	Validation set	
29 × 29	7396	200	120	80	6255001
43 × 43	3364				
59 × 59	1764				

Table C3.3.2. Training and test samples for SVM-RBF-BBIA-based urban land use classification (block 43×43)

Class ID	Training land use classes	Tile1	
		Number of training blocks	Number of test pixels
1	Residential	60	2717828
2	Commercial	15	184944
3	Industrial	35	845305
4	Transportation	22	359030
5	Urban green	36	1099990
6	Vacant land	20	873329
7	Water body/shadow	12	174575
Total		200	6255001

Table C3.3.3. Training and test samples for SVM-RBF-BBIA-based urban land use classification (block 59×59)

Class ID	Training land use classes	Tile1	
		Number of training blocks	Number of test pixels
1	Residential	70	2717828
2	Commercial	15	184944
3	Industrial	38	845305
4	Transportation	21	359030
5	Urban green	31	1099990
6	Vacant land	20	873329
7	Water body/shadow	05	174575
Total		200	6255001

C4. Features selection

Table C4.1. Features selection for SVM-RBF-OBIA-based urban land cover classification using SFS-HSIC

Best features	10	15	20	25	120
Name of features	10, 90, 4, 5, 6, 9, 3, 7, 8, 2	10, 90, 109, 68, 4, 5, 67, 3, 6, 9, 82, 2, 58, 84, 1	10, 107, 90, 68, 4, 5, 67, 6, 3, 9, 7, 84, 8, 2, 44, 92, 43, 91, 83, 82	10, 90, 109, 4, 5, 68, 6, 67, 9, 3, 7, 8, 2, 84, 1, 82, 43, 91, 83, 44, 92, 31, 20, 36, 100	1 to 120
Best sigma	1.29155	1.29155	1.29155	1.29155	-
HSIC	0.03345184	0.03670414	0.03734782	0.03969606	-
Over all accuracy	76.81	76.00	76.84	77.98	76.11

Table C4.2. Features selection for SVM-RBF-OBIA-based urban land use classification using SFS-HSIC

MRS Level	Best features	10	15	20	25	125
Level 1	Name of features	124, 125, 122, 10, 8, 7, 6, 123, 5, 4	124, 125, 122, 10, 123, 90, 84, 8, 83, 4, 7, 5, 6, 3, 67	124, 125, 122, 10, 8, 84, 7, 123, 6, 5, 4, 83, 67, 9, 3, 2, 87, 1, 86, 56	-	1 to 125
	HSIC	0.05350289	0.05858687	0.05503954	-	-
	Best sigma	0.4641589	0.4641589	0.4641589	-	-
	Over all accuracy	68.09	69.53	68.94		61.59
Level 2	Name of features	124, 125, 122, 10, 123, 90, 3, 4, 5, 6	124, 125, 122, 10, 123, 90, 3, 84, 4, 5, 6, 9, 82, 83, 8	124, 125, 122, 10, 123, 90, 107, 3, 4, 82, 5, 84, 119, 6, 83, 9, 2, 1, 67, 7	124, 125, 122, 10, 90, 123, 82, 3, 4, 84, 5, 83, 6, 9, 2, 76, 79, 1, 7, 80, 8, 119, 78, 75, 87	1 to 125
	HSIC	0.05497107	0.04976257	0.04873324	0.04559597	-
	Best sigma	0.4641589	0.4641589	0.4641589	0.4641589	-
	Over all accuracy	65.75	65.25	68.10	67.61	66.11
Level 3	Name of features	122, 125, 124, 10, 109, 110, 58, 107, 90, 113	125, 122, 124, 107, 10, 110, 109, 113, 84, 3, 90, 76, 4, 108, 119	24, 122, 125, 10, 107, 109, 110, 58, 114, 90, 84, 108, 3, 4, 83, 5, 113, 82, 123, 6	-	1 to 125
	HSIC	0.02967345	0.02559794	0.03132948	-	-
	Best sigma	0.4641589	0.4641589	0.4641589	-	-
	Over all accuracy	66.15	68.67	65.68	-	67.91

Note: Name and sequence of the 125 features explained the table C4.1 and C4.2

Name of the features	Sequence of the features	Name of the features	Sequence of the features
Meanband	1-8	Elliptical fit	109
Mean brightness	9	Rectangular fit	110
Mean NDVI	10	Roundness	111
GLCM (mean)	11-18	Main direction	112
GLCM (SD)	19-26	Shape Index	113
GLCM (homogeneity)	27-34	Boarder Index	114
GLCM (contrast)	35-42	Radius of largest enclosed ellipse	115
GLCM (dissimilarity)	43-50	Radius of smallest enclosed ellipse	116
GLCM (entropy)	51-58	Border length	117
GLCM(Angular second moment)	59-66	Width	118
GLCM (correlation)	67-74	Length-width ratio	119
GLDV(Angular second moment)	75-82	Number of pixels	120
GLDV(entropy)	83-90	Path density	121
GLDV(mean)	91-98	Aggregation index	122
GLDV(contrast)	99-106	Fractal dimension	123
Asymmetry	107	Cohesion	124
Compactness	107	Largest patch Index	125

Table C4.3. Features selection for SVM-RBF-BBIA-based urban land use classification using SFS-HSIC

Blocks	Best features	10	15	20	25	121
29 ×29	Name of features	120, 118, 121, 3, 4, 10, 5, 1, 2, 9	118, 120, 121, 3, 4, 10, 5, 1, 2, 9, 6, 58, 7, 89, 8	118, 120, 121, 10, 3, 4, 1, 58, 5, 2, 89, 9, 6, 8, 86, 7, 85, 50, 91, 87	118, 120, 121, 4, 5, 3, 10, 1, 2, 9, 6, 58, 116, 7, 115, 8, 89, 49, 92, 86, 85, 47, 94, 46, 95	1 to 121
	HSIC	0.06161469	0.05742712	0.0538051	0.06721506	-
	Best sigma	0.4641589	0.4641589	0.4641589	0.4641589	-
	Over all accuracy	67.49	67.70	63.28	63.34	62.43
43 ×43	Name of features	118, 121, 120, 5, 4, 58, 3, 10, 115, 116	118, 120, 121, 5, 10, 4, 115, 58, 116, 3, 2, 9, 1, 6, 89	118, 120, 121, 10, 116, 5, 115, 58, 4, 3, 2, 1, 89, 9, 6, 8, 52, 7, 75, 51	118, 120, 121, 10, 5, 58, 115, 116, 4, 3, 2, 1, 9, 8, 7, 6, 89, 51, 56, 75, 66, 26, 53, 87, 52	1 to 121
	HSIC	0.04928167	0.05237353	0.05753959	0.04798417	-
	Best sigma	0.4641589	0.4641589	0.4641589	0.4641589	-
	Over all accuracy	63.10	64.54	65.97	64.71	61.21
59 ×59	Name of features	118, 116, 115, 120, 58, 10, 5, 3, 4, 2	118, 120, 116, 115, 58, 3, 2, 4, 5, 10, 1, 9, 89, 6, 54	118, 116, 115, 121, 58, 3, 2, 4, 5, 1, 10, 54, 120, 9, 6, 53, 89, 7, 8, 55	118, 116, 115, 120, 3, 58, 5, 4, 10, 2, 1, 9, 54, 6, 52, 89, 7, 8, 53, 90, 84, 83, 80, 66, 29	1 to 121
	HSIC	0.04997003	0.05041109	0.04922437	0.04808981	-
	Best sigma	0.4641589	0.4641589	0.4641589	0.4641589	-
	Over all accuracy	60.78	58.53	58.31	58.07	55.27

Note: Name and sequence of the 121 features explained the table C4.3

Name of the features	Sequence of the features	Name of the features	Sequence of the features
Meanband	1-8	GLDV(Angular second moment)	75-82
Mean brightness	9	GLDV(entropy)	83-90
Mean NDVI	10	GLDV(mean)	91-98
GLCM (mean)	11-18	GLDV(contrast)	99-106
GLCM (SD)	19-26	LBP feature	107-114
GLCM (homogeneity)	27-34	MPPR feature	115-116
GLCM (contrast)	35-42	Path density	117
GLCM (dissimilarity)	43-50	Aggregation index	118
GLCM (entropy)	51-58	Fractal dimension	119
GLCM(Angular second moment)	59-66	Cohesion	120
GLCM (correlation)	67-74	Largest patch Index	121

C5. Parameter tuning

C5.1. Parameter tuning for SVM-RBF

Table C5.1.1. Best parameter for SVM-RBF-OBIA-based urban land cover classification

SVM-RBF parameter	MRS level 1, best 25 features
Name of features	10, 90, 109, 4, 5, 68, 6, 67, 9, 3, 7, 8, 2, 84, 1, 82, 43, 91, 83, 44, 92, 31, 20, 36, 100
Best gamma	0.166810
Best Cost	1000.00
Over all accuracy	78.21

Table C5.1.2. Experiment with size of training samples for SVM-RBF-OBIA-based urban land cover classification (tile1)

Size of training samples	Overall accuracy
50	71.00
100	75.16
150	77.29
200	**78.21**

Table C5.1.3. Best parameter for SVM-RBF-OBIA-based urban land use classification

SVM-RBF parameter	MRS level 1 best 15 features	MRS level 2 best 20 features	MRS level 3 best 15 features
Name of features	124, 125, 122, 10, 123, 90, 84, 8, 83, 4, 7, 5, 6, 3, 67	124, 125, 122, 10, 123, 90, 107, 3, 4, 82, 5, 84, 119, 6, 83, 9, 2, 1, 67, 7	125, 122, 124, 107, 10, 110, 109, 113, 84, 3, 90, 76, 4, 108, 119
Best gamma	0.05994843	0.1668101	0.4641589
Best Cost	1000	129.155	359.3814
Over all accuracy	70.58	68.10	69.08

Table C5.1.4. Best parameter for SVM-RBF-BBIA-based urban land use classification

Best features	Block 29 ×29 best 15 features	Block 43 × 43 best 20 features	Block 59 ×59 best 10 features
Name of features	118, 120, 121, 3, 4, 10, 5, 1, 2, 9, 6, 58, 7, 89, 8	118, 120, 121, 10, 116, 5, 115, 58, 4, 3, 2, 1, 89, 9, 6, 8, 52, 7, 75, 51	118, 116, 115, 120, 58, 10, 5, 3, 4, 2
Best gamma	1.29155	0.4641589	1.29155
Best Cost	16.68101	16.68101	16.68101
Over all accuracy	68.37	67.46	60.92

C5.2. Parameter tuning for CNN

C5.2.1. Tuning of learning and regularization parameters

Table C5.2.1.1. Learning and regularization parameters

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.01, 0.001
Weight decay, λ	0.01, 0.001, 0.0001
Dropout rate, d_r in (D1 & D2)	(0, 0.5), (0.25, 0.5) (0.5, 0.5)

Table C5.2.1.2. Fixed CNN configuration for selecting best learning and regularization parameters

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in $C_{1\text{-}2}$ and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C= convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

Table C5.2.1.3. 6-fold cross-validation of learning and regularization parameters (training samples 1000)

Experiment step	K-fold	Learning rate	Weight decay	Dropout, D1, D2	Overall accuracy (%)
1	1	0.01	0.01	(0, 0.5)	42.9850
2		0.01	0.001		47.9757
3		0.01	0.0001		43.4501
4	2	0.001	0.01		60.1607
5		0.001	0.001		58.6014
6		0.001	0.0001		54.7037
7	3	0.01	0.01	(0.25, 0.5)	50.2367
8		0.01	0.001		50.2389
9		0.01	0.0001		52.1664
10	4	**0.001**	**0.01**		**65.1242**
11		0.001	0.001		61.6627
12		0.001	0.0001		60.9216
13	5	0.01	0.01	(0.5, 0.5)	30.5377
14		0.01	0.001		22.5377
15		0.01	0.0001		43.4505
16	6	0.001	0.01		62.1612
17		0.001	0.001		57.3533
18		0.001	0.0001		59.2925

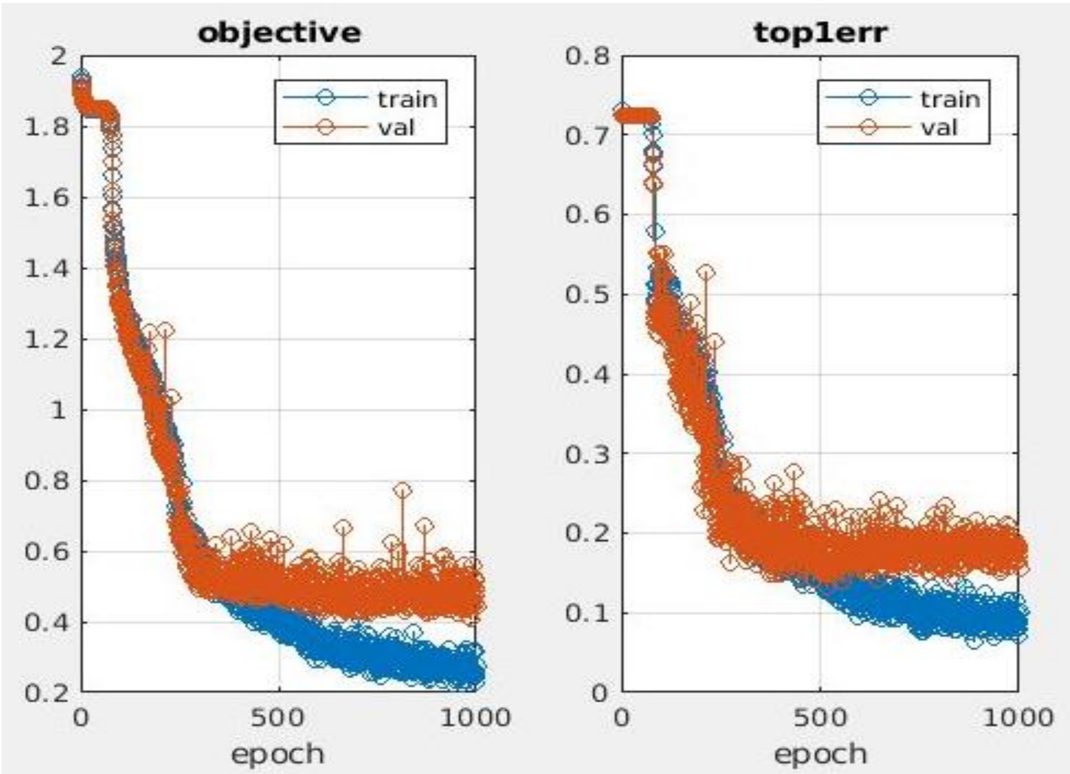


Figure C5.2.1.1. Shows objective loss and top layer error decreasing with increasing of epoch

C5.2.2. Tuning of CNN hyperparameters

Experiment on number of filters

Table C5.2.2.1. Best learning and regularization parameters

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.001
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0.25, 0.5)

Table C5.2.2.2. Number of filters experiment: CNN configuration

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in C_{1-2} and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8,16,32,64
Size of filters	5
Pooling size	2

Note: I=input, C= convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

Table C5.2.2.3. Overall accuracy varied with the number of filters

Number of filters	Overall accuracy (%)
8	**65.12**
16	61.30
32	61.82
64	62.55

Experiment on filter size

Table C5.2.2.4. Best learning and regularization parameters

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.001
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0.25, 0.5)

Table C5.2.2.5. Filter size experiment: CNN configuration

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in $C_{1\text{-}2}$ and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5,15, 25
Pooling size	2

Note: I=input, C= convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

Table C5.2.2.6. Overall accuracy varied with the number of filters

Size of filters	Overall accuracy (%)
5	**65.12**
15	61.86
25	62.58

Experiment on different convolutional layers with fixed FC=1

Table C5.2.2.7. Best learning and regularization parameters

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.001
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0.25, 0.5)

Table C5.2.2.8. Experiment on C_n layers with FC_1: CNN configuration

Hyperparameters	Value
Layers	I-C_n-A-P-D1—FC_1-A-D2—O-S-CP
Non-linearity (A=RELU) used in $C_{1\text{-}2}$ and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

Table C5.2.2.9. Overall accuracy varied with the number convolutional layers (C_n)

Convolutional layers	Overall accuracy (%)
2	**65.12**
3	43.45
4	43.45

Experiment on different fully-connected layers with fixed Convolutional layers=1

Table C5.2.2.10. Best learning and regularization parameters

Hyperparameters	Value
Batch size	10
Maximum number of epoch, τ	1000
Momentum, α	0.90
Learning rate, ϵ	0.001
Weight decay, λ	0.01
Dropout rate, d_r in (D1 & D2)	(0.25, 0.5)

Table C5.2.2.11. Experiment on FC_n layers with C_1: CNN configuration

Hyperparameters	Value
Layers	I-C_1-A-P-D1—C_2-A-P-D1—FC_n-A-D2—O-S-CP
Non-linearity (A=RELU) used in $C_{1\text{-}2}$ and FC_1	RELU
Non-linearity (S= SoftMax) used in O	SoftMax
Width of FC	128
Patch size	29
Number of filters, K	8
Size of filters	5
Pooling size	2

Note: I=input, C=convolutional layer, A =activation, P=max pooling, D=dropout, FC=fully connected layer, O=output, S= SoftMax, CP=class probability

Table C5.2.2.12. Overall accuracy varied with the number fully-connected layers (FC_n)

FC layers	Overall accuracy (%)
1	**65.12**
2	43.45
3	43.45

C6. SVM-RBF-OBIA-based Urban land cover classification

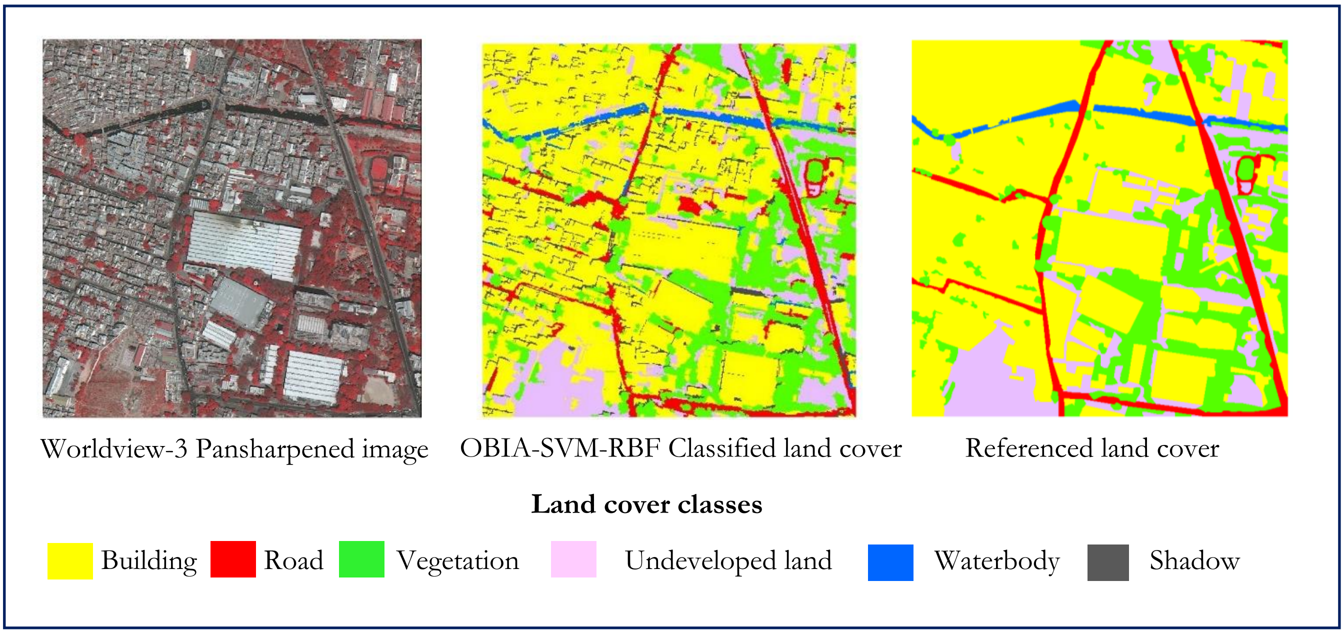


Figure C6.1. Shows SVM-RBF-OBIA-based classified urban land cover map of tile one

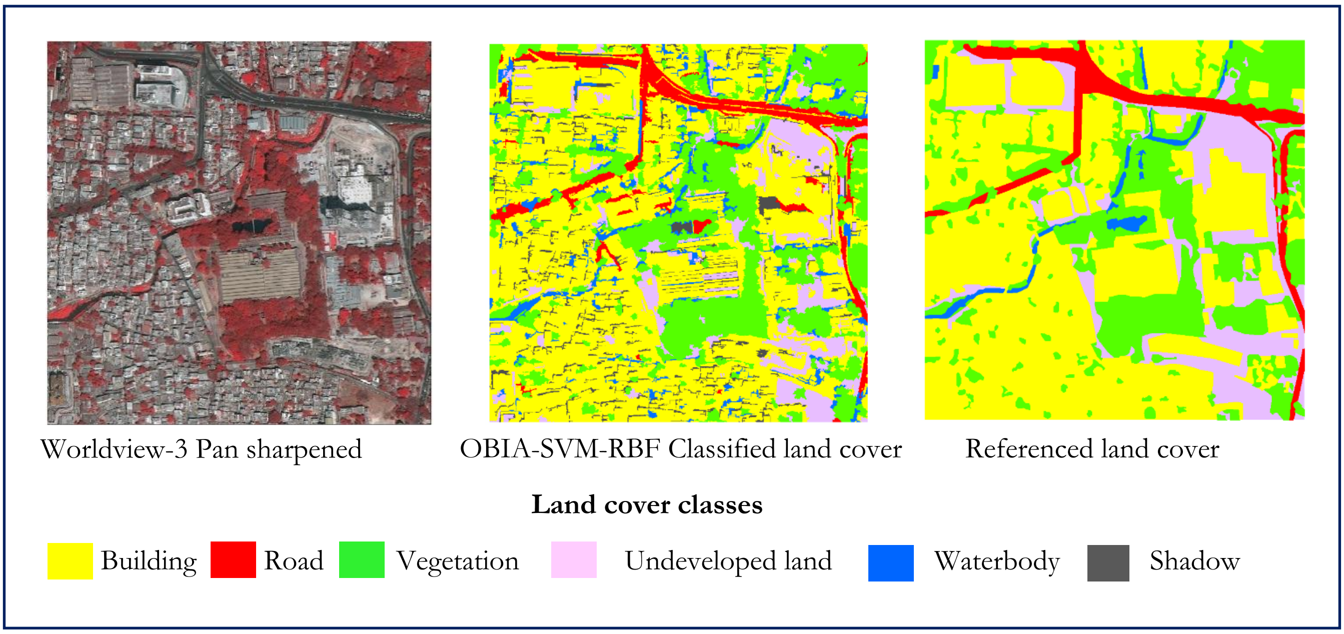


Figure C6.2. Shows SVM-RBF-OBIA-based classified urban land cover map of tile two

Table C6.1. Confusion matrix of SVM-RBF-OBIA-based urban land cover classification (tile one)

Land cover classes		Referenced						
		Building	Road	Vegetation	Undeveloped land	Water body	Shadow	UA %
Predicted	Building	3046055	58623	61721	188914	9103	18503	90.04
	Road	86856	245789	27624	28965	1900	6393	61.83
	Urban green	139098	14403	938103	133448	3353	2892	76.19
	Undeveloped land	222692	32797	78084	514625	2471	1167	60.41
	Water body	8317	2470	3124	5872	79519	4851	76.35
	Shadow	156369	15626	20695	23647	3060	67872	23.63
	PA %	83.24	66.48	83.07	57.47	80.00	0.66.75	

Over all accuracy =78.21%

Table C6.2. Confusion matrix of SVM-RBF-OBIA-based urban land cover classification (tile two)

Land cover classes		Referenced						
		Building	Road	Vegetation	Undeveloped land	Water body	Shadow	UA %
Predicted	Building	2776617	60139	135375	366442	6253	7935	82.82
	Road	30081	241080	25434	19563	12165	24606	68.31
	Urban green	86346	9259	1273112	49757	9146	36	89.17
	Undeveloped land	126466	5078	76182	356275	1528	38	62.99
	Water body	109235	19999	107787	15305	49640	5976	16.12
	Shadow	178743	8795	7287	19388	6883	27050	10.90
	PA %	83.95	70.01	78.34	43.09	57.98	41.21	

Over all accuracy =75.52%

C7. SVM-RBF-OBIA-based Urban land use classification

Sampled domain

Table C7.1. Confusion matrix of SVM-RBF-OBIA-based urban land use classification (tile one)

Land use classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	2064898	84444	220880	56649	31166	109510	14480	79.97
	Commercial	8071	31256	3960	7252	3468	25152	152	39.41
	Industrial	107380	18975	511356	726	6744	18107	844	77.00
	Transportation	64611	12730	30885	227042	31279	56683	1179	53.50
	Urban green	125381	2453	27076	16553	929355	120139	5666	75.77
	Vacant land	194488	32095	38969	35589	70099	516325	2364	58.02
	Waterbody/ shadow	152999	2991	12179	15219	27879	27413	149890	38.57
	PA %	75.98	16.90	60.49	63.24	84.49	59.12	85.86	

Over all accuracy =70.83% and kappa= 0.6089

Unsampled domain

Table C7.2. Confusion matrix of SVM-RBF-OBIA-based urban land use classification (tile two)

Land cover classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	1764285	184770	257582	44468	65112	312700	5274	66.98
	Commercial	111213	72283	105839	13074	60372	40986	10034	17.47
	Industrial	60344	121845	43580	8013	1082	26827	1351	16.57
	Transportation	80949	17649	51524	255846	44001	54877	51295	46.00
	Urban green	126713	10887	12101	9283	1265546	43045	12098	85.53
	Vacant land	122278	23618	20221	2081	120721	311475	3410	51.59
	Waterbody/ shadow	170452	4028	9436	8195	31886	21643	58709	19.29
	PA %	72.42	16.61	8.71	75.04	79.66	38.38	41.29	

Over all accuracy =60.30% and kappa=0.4703

C8. SVM-RBF-BBIA-based Urban land use classification

Sampled domain

Table C8.1. Confusion matrix of SVM-RBF-BBIA-based urban land use classification (tile one)

Land cover classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	2382473	96301	327387	80733	115331	249111	58256	71.99
	Commercial	20679	46573	15635	885	29940	28061	735	32.68
	Industrial	83067	20064	437063	64	4521	12672	1181	78.24
	Transportation	125190	9995	34933	220491	89328	84030	43326	36.31
	Urban green	33601	2755	12003	16664	709502	64919	6347	83.89
	Vacant land	66176	9232	16344	23314	123598	427187	2604	63.91
	Waterbody/ shadow	6642	24	1940	16879	27770	7349	62126	50.62
	PA %	87.66	25.18	51.70	61.41	64.50	48.91	35.59	

Over all accuracy =68.51% and kappa=0.5581

Unsampled domain

Table C8.2. Confusion matrix of SVM-RBF-BBIA-based urban land use classification (tile two)

Land cover classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	2043710	315612	351833	100746	702214	534894	24602	50.17
	Commercial	8611	25706	2605	870	20571	9765	7852	33.83
	Industrial	9198	16980	13791	1161	796	7858	67	27.66
	Transportation	88067	16031	22752	187270	407581	60614	48075	22.55
	Urban green	7239	622	1230	7	293836	4618	6057	93.70
	Vacant land	277780	59262	107698	4947	139827	189551	8954	24.05
	Waterbody/ shadow	1629	867	374	45959	23895	4253	46564	37.69
	PA %	83.89	5.91	2.76	54.92	18.50	23.36	32.75	

Over all accuracy = 44.77% and kappa=0.2200

C9. CNN-BBIA-based Urban land use classification

Sampled domain

Table C9.1. Confusion matrix of CNN-BBIA-based urban land use classification (tile one)

Land cover classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	1998924	84400	315337	55140	70558	192813	45730	72.35
	Commercial	117826	51021	41308	7779	8030	25468	1313	20.19
	Industrial	118322	18711	409434	1709	381	4254	3138	73.65
	Transportation	113948	5186	25909	224863	29493	45181	59772	44.58
	Urban green	31187	1174	4368	18971	809760	80260	4812	85.19
	Vacant land	335339	24431	48599	24930	167853	520572	862	46.37
	Waterbody/ shadow	2282	21	350	25638	13915	4781	58948	55.65
	PA %	73.55	27.59	48.44	62.63	73.62	59.61	33.77	

Over all accuracy =65.12% and kappa=0.5274

Unsampled domain

Table C9.2. Confusion matrix of CNN-BBIA-based urban land use classification (tile two)

Land cover classes		Referenced							
		Residential	Commercial	Industrial	Transportation	Urban green	Vacant land	Water body/ shadow	UA %
Predicted	Residential	1956578	277392	253054	91809	140711	434625	26976	61.51
	Commercial	79931	20210	25939	3339	75720	83219	1295	6.98
	Industrial	215319	111468	57753	7374	356	30043	1751	13.62
	Transportation	32510	5397	1787	226186	72890	16757	59511	54.50
	Urban green	1661	0	155	746	571591	1051	6196	98.31
	Vacant land	150059	20570	161534	4358	723117	245746	4161	18.77
	Waterbody/ shadow	176	43	61	7148	4335	112	42281	78.07
	PA %	80.31	4.65	11.54	66.34	35.98	30.28	29.74	

Over all accuracy =49.89% and kappa=0.3216

B11. Performance measurement and evaluation

Sampled domain

Table C10.1. Performance of selected machine learning algorithms for urban land use classification (tile one)

Land use classes	SVM-RBF- OBIA			SVM- RBF-BBIA			CNN-BBIA		
	Recall	Precision	F1-score	Recall	Precision	F1-score	Recall	Precision	F1-score
Residen.	79.97	75.98	77.92	71.99	87.66	79.05	72.35	73.55	72.94
Commer.	39.41	16.90	23.66	32.68	25.18	28.45	20.19	27.59	23.31
Industri.	77.00	60.49	67.75	78.24	51.70	62.26	73.65	48.44	58.44
Trans.	53.50	63.24	57.96	36.31	61.41	45.64	44.58	62.63	52.09
Ur.green.	75.77	84.49	79.89	83.89	64.50	72.93	85.19	73.62	78.98
Vac.land	58.02	59.12	58.56	63.91	48.91	55.41	46.37	59.61	52.16
Water/ shadow	38.57	85.86	53.23	50.62	35.59	41.79	55.65	33.77	42.03
Overall	60.32	63.73	59.85	59.66	53.57	55.08	56.85	54.17	54.28

Unsampled domain

Table C10.2. Performance of selected machine learning algorithms for urban land use classification (tile two)

Land use classes	SVM-RBF-OBIA			SVM- RBF-BBIA			CNN-BBIA		
	Recall	Precision	F1-score	Recall	Precision	F1-score	Recall	Precision	F1-score
Residen.	66.98	72.42	69.59	50.17	83.89	62.79	61.51	80.31	69.66
Commer.	17.47	16.61	17.03	33.83	5.91	10.06	6.98	4.65	5.58
Industri.	16.57	8.71	11.42	27.66	2.76	5.01	13.62	11.54	12.50
Trans.	46.00	75.04	57.04	22.55	54.92	31.98	54.50	66.34	59.84
Ur.green.	85.53	79.66	82.49	93.70	18.50	30.89	98.31	35.98	52.68
Vac.land	51.59	38.38	44.01	24.05	23.36	23.70	18.77	30.28	23.17
Water/ shadow	19.29	41.29	26.30	37.69	32.75	35.05	78.07	29.74	43.07
	43.35	47.44	43.98	41.38	31.73	28.50	47.39	36.98	38.07

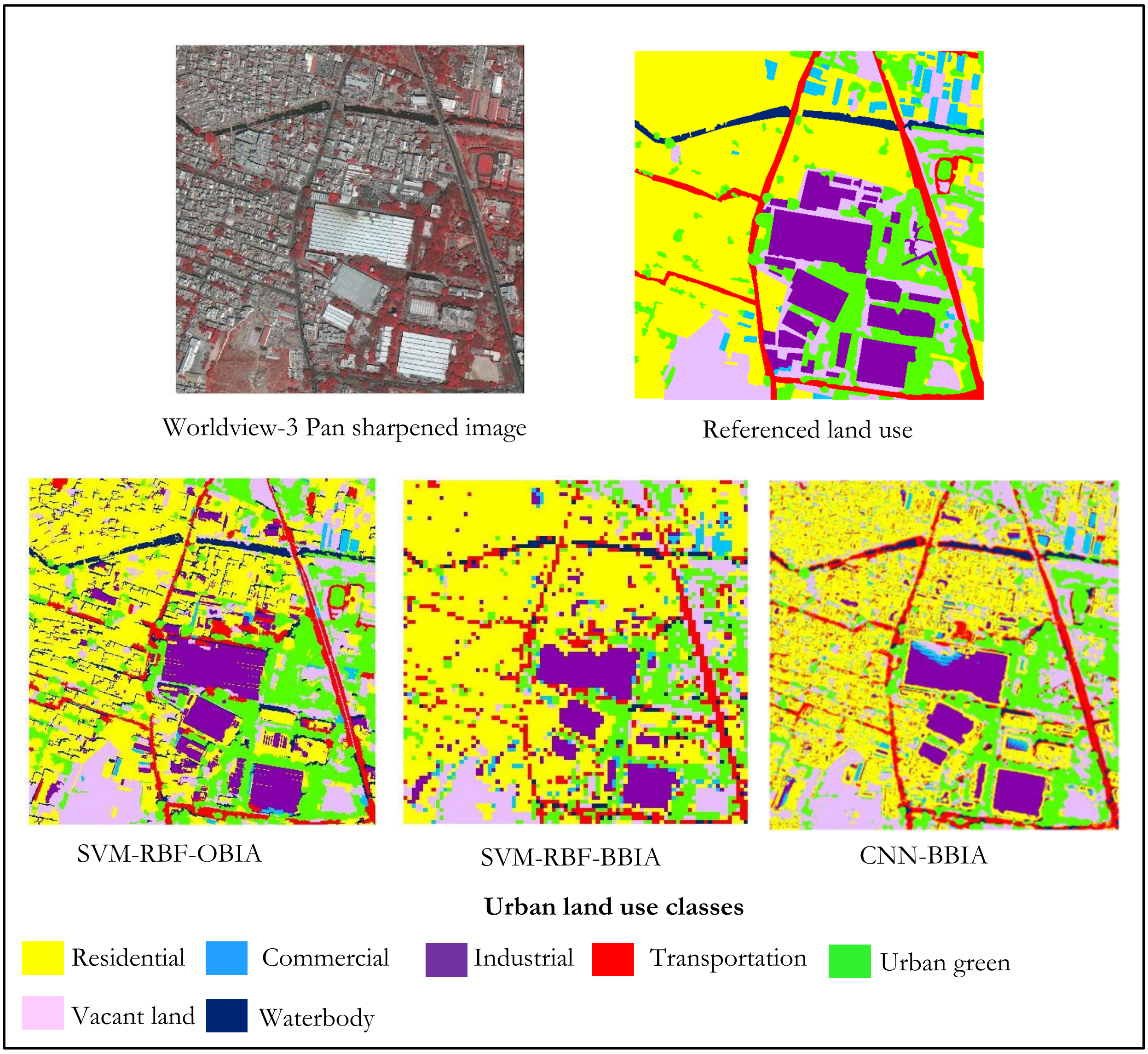


Figure C10.1. Compares quality of classified map with referenced land use map (tile one sampled domain)

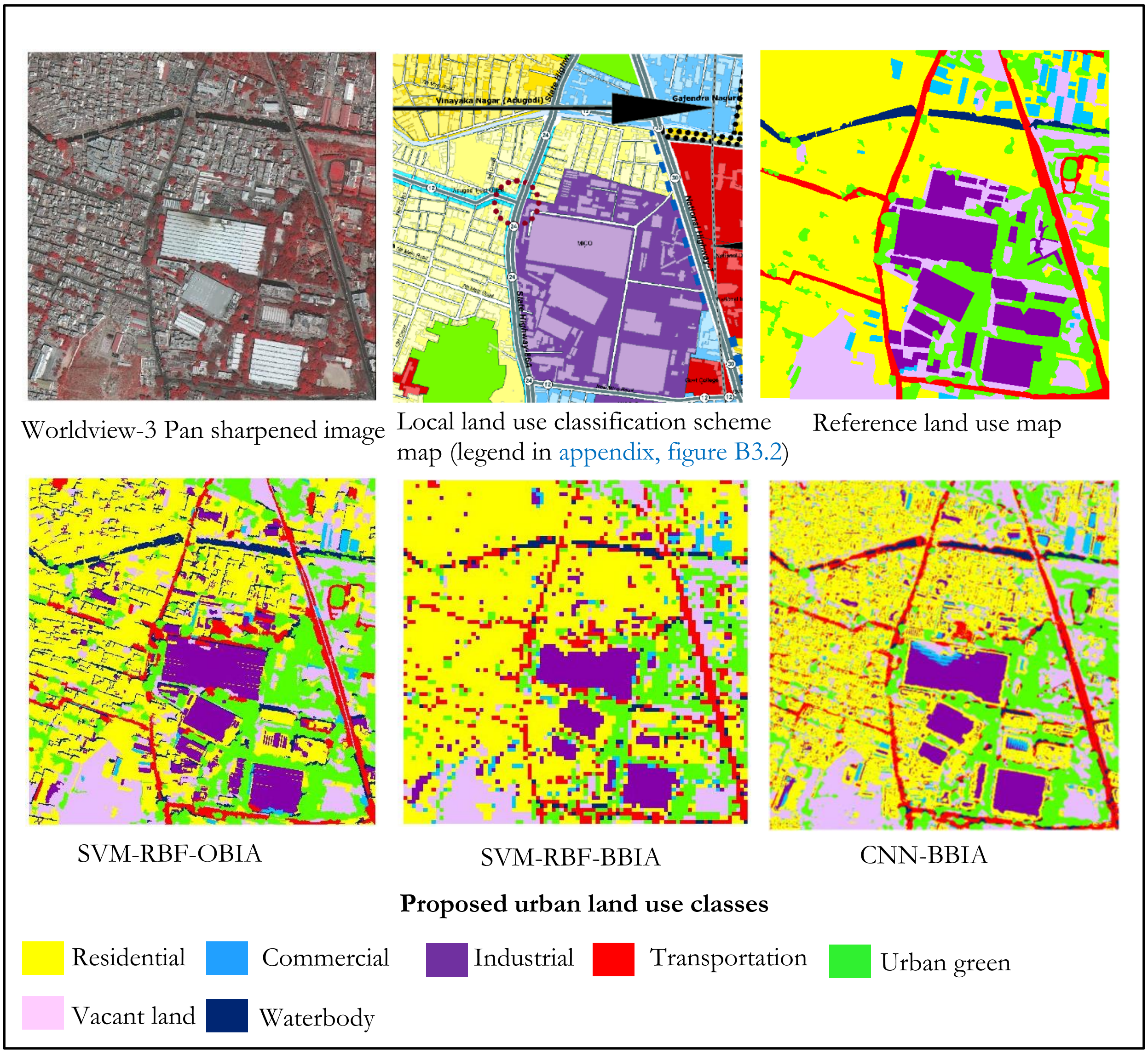


Figure C10.2. Compares quality of proposed land use map with local land use classification scheme map (tile one sampled domain)