

Lamyaa Gamal El-deen Taha¹, Rania Elsayed Ibrahim²

A Machine Learning Model for Improving Building Detection in Informal Areas: A Case Study of Greater Cairo

Abstract: Building detection in Ashwa'iyat is a fundamental yet challenging problem, mainly because it requires the correct recovery of building footprints from images with high-object density and scene complexity. A classification model was proposed to integrate spectral, height and textural features. It was developed for the automatic detection of the rectangular, irregular structure and quite small size buildings or buildings which are close to each other but not adjoined. It is intended to improve the precision with which buildings are classified using scikit learn Python libraries and QGIS. WorldView-2 and Spot-5 imagery were combined using three image fusion techniques. The Grey-Level Co-occurrence Matrix was applied to determine which attributes are important in detecting and extracting buildings. The Normalized Digital Surface Model was also generated with 0.5-m resolution. The results demonstrated that when textural features of colour images were introduced as classifier input, the overall accuracy was improved in most cases. The results show that the proposed model was more accurate and efficient than the state-of-the-art methods and can be used effectively to extract the boundaries of small size buildings. The use of a classifier ensemble is recommended for the extraction of buildings.

Keywords: multi-source image fusion, random forest, support vector machine, DEM extraction, unplanned unsafe areas, remote sensing

Received: 22 July 2021; accepted: 26 November 2021

© 2022 Authors. This is an open access publication, which can be used, distributed and reproduced in any medium according to the Creative Commons CC-BY 4.0 License.

¹ National Authority for Remote Sensing and Space Sciences, Cairo, Egypt, email: lamyaa@narss.sci.eg,  <https://orcid.org/0000-0003-2890-2498>

² National Authority for Remote Sensing and Space Sciences, Cairo, Egypt, email: raniaelsayed@narss.sci.eg,  <http://orcid.org/0000-0001-5116-5866>

1. Introduction

Unsafe area and unplanned areas are an umbrella concept of informal settlement. Within the Egyptian context, they have been known as 'Ashwa'yyat'. Unsafe areas are identified by the presence of a life-threatening threat, whereas unplanned areas are identified by noncompliance with planning and construction laws [1]. They are suffering from problems of narrow streets, air pollution etc.

As a result, the Egyptian government developed a national strategy for improving Ashwa'yyat, followed by an action plan to establish new towns in the desert toward "Egypt Without Ashwa'yyat" by 2030 to achieve Sustainable Development Goal 11; "Make cities inclusive, safe, resilient, and sustainable" [2]. Building boundaries are traditionally drawn using photogrammetric stereo plotters and hand digitization from digital photos in stereo vision. However, this is a time-consuming and exhausting operation that necessitates the use of trained personnel and expensive equipment. As a result, utilising automated ways to extract buildings has a lot of potential and value [3]. Many complex factors that influence buildings extraction from satellite images, such as scene complexity, building inconstancy, and image resolution, affect overall accuracy for building extraction and detection [4]. The main challenge of this study is to provide precise and high accurate framework for buildings detection that is essential for a variety of applications [5], including urban planning and management, infrastructure development, and so on.

The fundamental problem with building extraction methods is that the building class is confused with other object classes such as shadows, vegetation, and the ground. Other issues with misclassification include misclassifying a non-building as a building and mixing trees and shadows. These misclassification issues, which are caused by a single dataset and approach, have a negative impact on the accuracy of the classification process. As a result, various methods and strategies have been proposed to address the issues raised by the complexity of the classification process [6]. Another issue is that the building's roof could be made of several surface materials with varying reflecting qualities, which is a challenge.

As a result of these issues, the computer vision field has become increasingly complicated, resulting in errors and complexity in automatic building detection. Therefore, many algorithms and strategies have been developed to solve these problems [7].

In order to extract buildings, various approaches were proposed by many researchers which include image-based [4, 8–11], LiDAR-based [12, 13] and data fusion-based methods [14, 15]. Although these approaches have shown promise in real-world applications include highly complicated scenes, such as the presence of barriers posed by surrounding objects, such as trees. To overcome these challenges, we proposed a framework to integrate multi-source image fusion technology with digital surface model (DSM) using machine learning (ML) techniques for buildings detection.

Sohn and Dowman [16] suggested an algorithmic method for extracting buildings from IKONOS photos in densely populated areas. They used huge detached buildings in their investigation without performing any accuracy analysis or modeling of the structure features.

Support vector machines (SVMs), ensemble classifiers (i.e. random forest (RF)), and deep learning algorithms are the most commonly used techniques in the remote sensing society [17]. Belgiu and Drăguț [18] employed the RF classifier to successfully map urban buildings. In addition, single-date MODIS data has been used to categorise urban impervious surfaces [19].

Several studies [20, 21] have investigated the potential of RF classification to improve urban classification from LiDAR data. In case of input datasets that could be used to improve categorization, it is critical to involve only related datasets to reduce computational burden without immolating accuracy [22]. In this case, an RF classifier was used to assess the achievement of each data source to the results [22, 23].

The SVMs algorithms are kernel-based learning algorithms used in various applications [17]. Sigmoid, polynomial, radial basis etc are kernel models used to build different SVMs [24]. The performance of SVMs is determined by the appropriate selection of a kernel function [25]. As a result, another goal of this research is to compare RF and SVM classification algorithms.

The process of combining images from multiple sources into a single imagery is referred to as multi-source image fusion technology, where the resulted fused image would be more beneficial than any of the input imageries, and it has major importance to the photogrammetry tasks in computer vision [26–29]. A detailed review can be found in [30].

Recently, few investigations have been done on the multi-source image fusion classification using machine learning. The authors of [18, 31] developed a pixel-wise classification technique, called machine multi-level fusion network, and the authors of [19, 32] developed fusion-based approach called ‘SubFus’ with capability to integrate remotely sensed data and ancillary dataset for land cover classification.

The current study developed a model to integrate multi-source image fusion technology with the normalized digital surface model (nDSM) and texture data using ML techniques for building detection and extraction. grey-level co-occurrence matrix (GLCM) was used to determine which attributes are important in detecting and extracting buildings. WorldView-2 and Spot-5 images were fused using three image fusion techniques: modified-IHS, wavelet-PCA, wavelet-IHS.

To classify the dataset into buildings and non-buildings, ML techniques such as SVMs and RF were used. To ensure the reliability of the results, dataset was divided into 70% for training and 30% for validation. The main objectives of this study are development of ML model for buildings detection and determine which attributes are important in detecting and extracting buildings. The performance of the RF is compared to that of SVMs. The results revealed that our approach outperforms other peer methods for building detection, indicating the effectiveness of our method.

2. Field of Study and Data Sources

The study area (Fig. 1) is part of Greater Cairo. Greater Cairo is one of the biggest urban areas in Middle East and Africa. The study area is located between 29°54'50"N and 31°16'14"E. It is area is approximately 30 km². The experimental tests were carried out in the residential urban blocks selected in Greater Cairo.

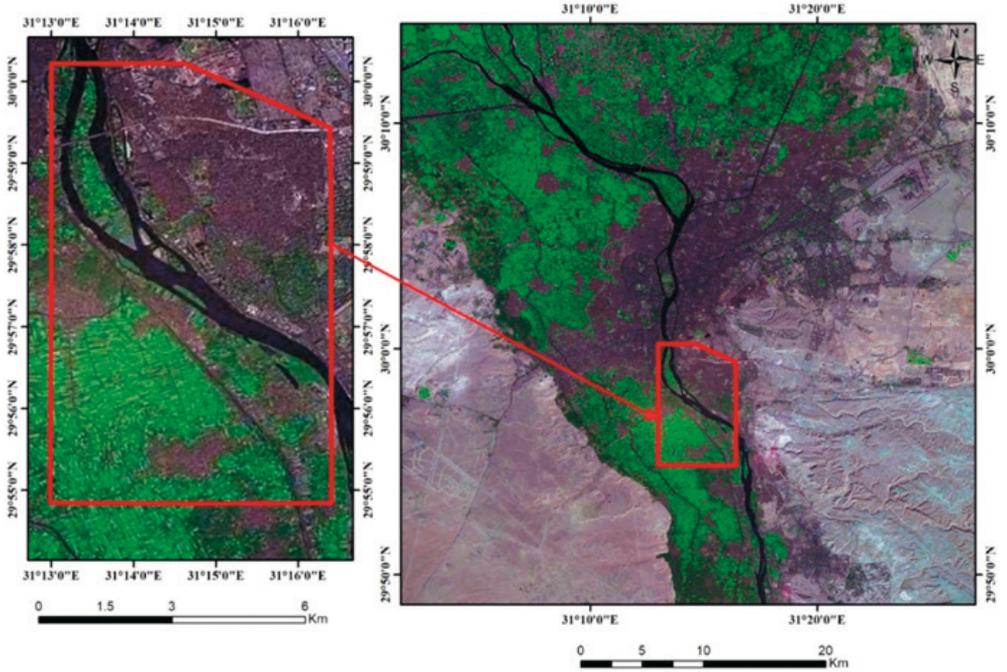


Fig. 1. Location map of study area

The data includes representative scene of a mix of low- and high-story buildings with a long range of rooftop structures and small size buildings. The following data sources are used.

- Cloud free WorldView-2 panchromatic stereo imagery (Tab. 1). Dated 12 Jan 2018, with 0.5-m resolution. The overlap of images is 90%. Worldview stereo scenes are supplied with rational polynomial functions (RPC) sensor model, derived from orbit and attitude information.

Table 1. WorldView-2 image characteristics

Band	Spectral range [μm]	Resolution [m]
Pan	0.45–0.8	0.46 GSD (0.52 at 20° off-nadir)

- Spot-5 image (Tab. 2). Dated 2 July 2009. The study area is a subset from the scene.

Table 2. Spot-5 image characteristics

Band	Spectral range [μm]	Resolution [m]
B1 (Green)	0.49–0.61	10
B2 (Red)	0.61–0.68	10
B3 (NIR)	0.78–0.89	10

- Thirty two DGPS ground control points and sixteen check points obtained from the field with 10-cm accuracy in X, Y, Z. The location of the points was selected using random stratified method to constitute various land cover classes of the area.

3. Methodology

Figure 2 illustrates the building detection framework and an illustration of the framework will be described in the following sections.

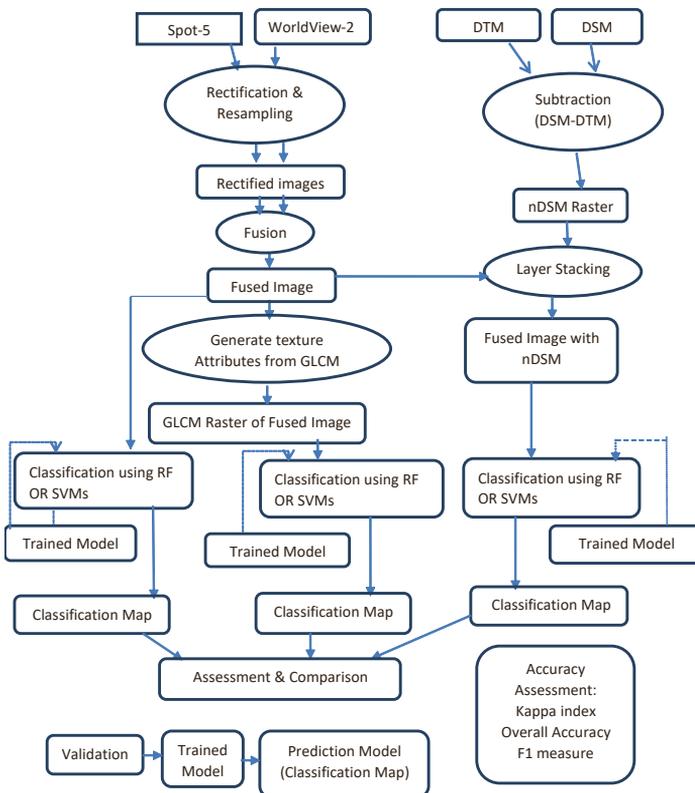


Fig. 2. Framework of building detection using machine learning algorithms

3.1. Image Orientation

The LPS environment [33] was used to produce Digital Surface Model (DSM) and image orientation of the stereo images. The process flow chart is illustrated in Figure 3.

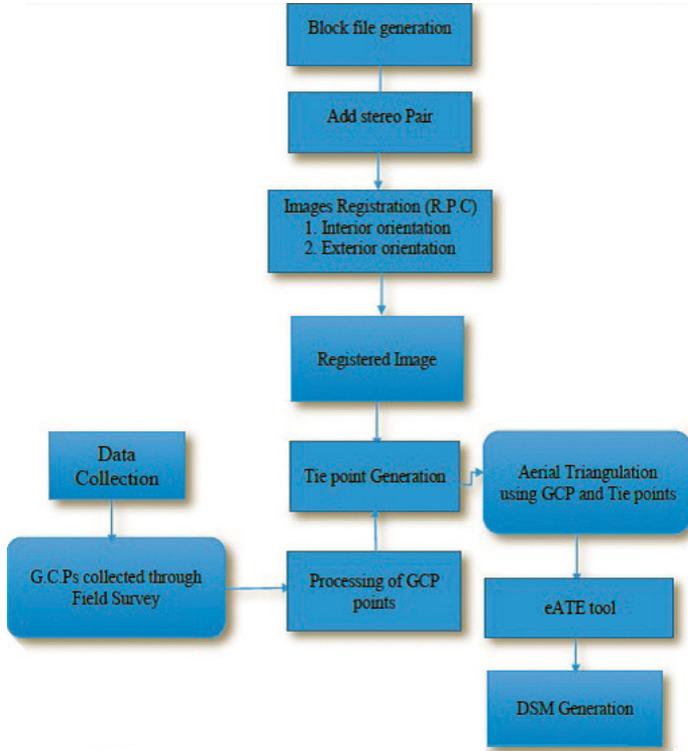


Fig. 3. Process flowchart

Creating Block (.blk) Files

The process in LPS starts with making a block project file defining the geometric model as RPC model. The RPC file includes the third degree polynomial coefficients that relate an image coordinates to its corresponding ground coordinates. UTM projection and WGS 84 datum have been assigned as the horizontal and vertical coordinates inside the block project.

Setting up Internal and External Orientation

The interior and exterior orientations of stereo pairs are recovered by extracting information from RPC file. Interior orientation sets the internal geometry of a sensor at the image capturing time while exterior orientation set the position and angular orientation of the sensor that captured the image [34].

Adding GCP Points and Automated Tie Point Generation

Automatic GCP and tie points collection were used to ensure the relative orientation link the images and get a stable block. It matches a point in one image, with its corresponding in the other (stereomate) image using a math model (rational function model). During Image Matching in LPS, a link window occurs between the reference image and the neighbouring overlapping image. Because most buildings in very high-resolution images appear similar, a larger value of moving window should be specified for more robust matching results. Total 150 tie points were generated in the overlapping area whose ground coordinates are not known. Using classical point measurement tool four ground points are added to the images. AeroTriangulation errors were 0.15 pixels with GCPs, and 0.731 pixels without GCPs.

DSM Generation

The triangulation process was run after adding GCPs and tie points to check the accuracy for GCPs and tie points. Thirty independent check points were utilized for DSM evaluation. The results revealed the RMS of 0.1 m was achieved using RPC.

Finally, the DSMs were produced with 0.5-m resolution. nDSM was produced by subtracting DTM from DSM. The DSM was generated using csf ground filtering of the DSM las point clouds. Consequently, Buildings_MinArea = 10 m², Buildings_MinWidth = 3 m, Trees_MinHeight = 1.3 m, Trees_MaxHeight = 50 m, Trees_MinRadius=2m, Trees_MaxRadius=6m. Output_Trees=0, Output_PowerLines=0.

Figure 4a shows resulted DSM of the study area, while Figure 4b shows normalized DSM (nDSM).

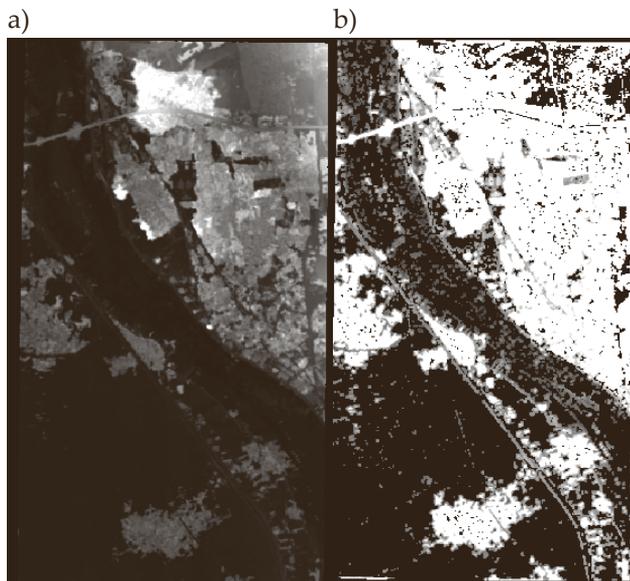


Fig. 4. Generation of digital surface models: a) DSM; b) nDSM

3.2. Image Ortho-rectification

The images were rectified using the DSMs with a pixel size 0.5 m and nearest neighbour resampling. The thirty DGPS check points were applied for orthoimage assessment resulted in RMS in X = 0.66, Y = 0.49, RMST = 0.82 using RPC only.

Rectification of Spot-5 Image

The Spot-5 image was rectified using 16 well-distributed GCPs obtained via DGPS. Images were projected to the UTM coordinate system using second order polynomials and the nearest neighbour algorithm as a resampling method during this process. The Spot-5 image was resampled to 0.5-m resolution to match the world view image. The precision was tested using 16 evenly distributed GCPs points. For the second order polynomial, the RMS error of check points was RMST = 0.321. Spot-5 was co-registered to the WorldView-2 image and resampled to a pixel size of 0.5 m using nearest neighbour.

3.3. Image Fusion

Three image fusion techniques (modified-IHS, wavelet-PCA, wavelet-IHS) have been applied on Spot-5 and WorldView-2 images. Color preservation and spatial improvement were statistically evaluated in all images.

Spectral Evaluation

In general, the quality of an image-fusion technique might be determined by comparing the fused image to a reference image. This comparison can be done both visually and statistically [35].

The quality evaluation parameters used for image fusion are: mean, standard deviation (Std), correlation coefficient and bias of mean [36–38].

3.4. Textural Feature Analysis Derived from Remote Sensing Data

Texture measures such as the co-occurrence matrix have shown their potential for enhancing classification performance in urban environments [29, 39].

In our study, GLCM was used to derive textural features using a small window of 3×3 pixels. ENVI was used for Extracting GLCM. The computed texture features were: mean, variance, homogeneity, correlation, contrast, second moment, dissimilarity, entropy [40].

3.5. Classification Using Machine Learning Algorithms

Various classification strategies are used in building detection for image classification such as maximum likelihood, support vector machines etc. [41]. In the current research, classifications were done using machine learning techniques; RF and SVMs. It is concerned with a binary classification problem. It divides the image into two categories: buildings and non-buildings. The classes are saved as

regions of interest (ROIs) in a shapefile. Each ROI has one target value (i.e., the class label) and several attributes (i.e. the features or observed variables). These techniques can train a classifier using ROIs, and then use the relationships discovered during the training process to classify the remaining pixels. The classification aims to produce a model that predicts the target values of the test data based solely on the test data attributes. All tests were conducted using the same training sets. The training data consists of 623 and 1,833 training pixels for buildings and non-buildings respectively. All pixels are distributed randomly. Figure 5 shows the distribution of pixels over the image. The classifier was trained with 30% of validation per class and 70% of training.

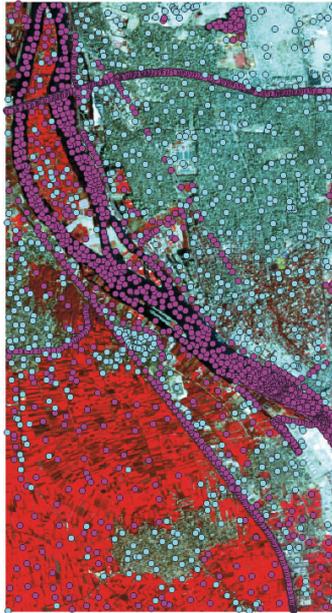


Fig. 5. Distribution of ROIs, light cyan pixels represent buildings and dark magenta pixels represent non-buildings

Classifications are implemented using nine schemes:

- 1) A file of fused image using wavelet-IHS only.
- 2) A composite file of fused image using wavelet-IHS and nDSM.
- 3) A composite file of GLCM raster of fused image using wavelet-IHS
- 4) A file of fused image using wavelet-PCA only.
- 5) A composite file of fused image using wavelet-PCA and nDSM.
- 6) A composite file of GLCM raster of fused image using wavelet-PCA
- 7) A file of fused image using modified-IHS only.
- 8) A composite file of fused image using modified-IHS and nDSM.
- 9) A composite file of GLCM raster of fused image using modified-HIS.

These files were used as predictor variables in the classification model. Classifications using all these schemes were repeated once for RF and a second time with SVMs, separately and individually.

Classification Using Random Forest

RF is a model that uses a random grid search to fit a figure of decision tree classifiers on different sub-samples of the dataset and then averages them to enhance predictive accuracy and control over-fitting. Classifiers are integrated in the scikit-learn implementation by averaging their probabilistic prediction [42]. Cross-validation was used to select the parameters that best generalised the data; find the best figure of trees, and the best figure of maximum features. The most important parameters are n estimators and max. features, where n estimators is the number of trees, and the higher the number, the better.

When dividing a node, max features is the size of the random subsets of features. Either all input features or a random subset of size max features are used to find the optimal split.

Classification Using SVMs

SVMs are a set of powerful models for regression, classification and outliers detection [40]. SVMs with Gaussian radial basis function (RBF) is used because it confirmed to be effective in remote sensing applications [43]. To find the optimal parameters, cross-validation was utilised. To train the dataset, the best parameters C and gamma were used. Finally, the model was put to the test.

The penalty parameter of the error term is parameter C . A high C means classifying all training examples correctly. Best C gave the best results on the hold out data.

The parameter gamma is the kernel coefficient. In calculating the separation line, a low value takes only nearby points while the high value takes all the data points. Best 'gamma' that gave the best results on the hold out data.

Radial basis function (RBF) kernel $K(x, y)$ is given by:

$$K(x, y) = \exp^{-\gamma \|x-y\|^2} \quad (1)$$

Evaluation of the Classification

The sklearn metrics module executes several score, and utility functions to evaluate classification performance including kappa index, overall accuracy (OA) and F1 measure. OA statistical index is a number of correct classified pixels/total number of pixels [44]. The kappa index is a discrete multivariate analysis tool for assessing the accuracy of classification maps. The kappa coefficient is derived from the error matrix and is used to evaluate how well the classification performs in comparison

to the reference data. The kappa statistic is used to see if a categorization based on remotely sensed imagery is better than random.

The difference between the real agreement (main diagonal total) and the chance agreement (row or column totals) of the matrix is the kappa coefficient of agreement.

Because it takes into account all elements of the confusion matrix, the kappa coefficient was recommended. The kappa can be defined as:

$$K = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \quad (2)$$

and it is computed as:

$$K = \frac{\sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{it} \cdot X_{ti})}{N \cdot N - \sum_{i=1}^r (X_{it} \cdot X_{ti})} \quad (3)$$

where:

- r – number of rows in error matrix,
- X_{ii} – number of observations in row (i) and column (i) on the major diagonal,
- X_{it} – total of observations in row (i) shown as marginal total of the matrix,
- X_{ti} – total of observations in column (i) shown as marginal total at the bottom of the matrix,
- N – total number of observations, included in the matrix [45].

The use of $F1$ (a combination of precision and recall) is a main difference between traditional RS and deep learning RS [46]. In most cases, neither precision nor recall can provide a comprehensive assessment of classification performance. Precision is the proportion of identified positives that are actually positive, whereas recall is the proportion of positives that are correctly identified. To compute the score, the $F1$ -score takes into account both the producer's accuracy (PA) and the user's accuracy (UA, also known as precision and recall) [47]:

$$F1 = \frac{UA \cdot PA}{UA + PA} \quad (4)$$

4. Results and Discussion

In this research a new innovative building detection model has been designed and developed which will reduce the system processing time and space complexity and give better recognition rate as compared to other existing techniques.

Spot-5 has been co-registered to the WorldView-2 image. The used resampling technique was nearest neighbour for resampling to the pixel size of 0.5 m. DEM was generated using ground filtering technique. Followed by calculating nDSM by subtracting DEM from DSM. Thirty check points were utilized for DSM evaluation. The resulted RMS was 0.1 m using RPC only and free of trees.

4.1. The Importance of Different Features

In this study, we hypothesized that derived attributes using GLCM matrix can be used to improve buildings detection. For GLCM construction and texture calculation, a small window of 3×3 pixels was used using ENVI 5.1 Software. Results revealed that GLCM variance and contrast were the best Haralick textural features for discriminating buildings and GLCM correlation was the least effective one.

4.2. An Assessment of Image Fusion

In this study, image fusion of multispectral Spot-5 image (10-m resolution) and WorldView-2 pan image (0.5-m resolution) has been achieved to improve the information appear in the images and to raise the reliability of the performance. Two images are co-registered and resampled to 0.5-m resolution.

Three image fusion techniques have been compared:

- wavelet-IHS,
- modified-IHS,
- wavelet-PCA.

Results show that image fusion enhanced the resolution. Three fused images gave better interpretation than multispectral image. Several indices were utilized to assess the performance of fusion. In this study, the spectral quality of fused images was assessed by:

- standard deviation,
- mean,
- correlation coefficient,
- bias of mean.

Table 3 gives quality metrics of fused images. The results revealed that the mean of the original image was 58.415. The average mean of all bands of fused images using wavelet-PCA, modified-IHS, wavelet-IHS were 58.148, 58.070 and 58.070 respectively. The standard deviation of the original image was 16.199 and the average standard deviation of all bands of fused image using wavelet-PCA, modified-IHS and wavelet-IHS were 16.111, 16.287 and 16.287 respectively. It is concluded that wavelet-PCA gave the least pixel errors and the three fused methods preserved the maximum sensible to the spectral quality of the original imagery. Correlation coefficient has been computed for the three fusion methods. The average correlation coefficient of all methods was 1.00 which indicates that the fused image is similar to the original MS image and no spatial information is put in the fused image from panchromatic image.

Table 3. Quality metrics of fused images

Bands	Mean of original image	St. dev. of original image	Mean of fused image	St. dev. of fused image	Diff. in mean	Diff. in st. dev.	Bias of mean	Fusion technique
Band 1	59.718	20.685	59.317	20.742	0.401	-0.057	0.007	Wavelet-PCA
Band 2	54.619	15.355	54.411	15.197	0.208	0.158	0.004	
Band 3	60.908	12.558	60.715	12.394	0.194	0.164	0.003	
AVERAGE	58.415	16.199	58.148	16.111	0.268	0.088	0.005	
Band 1	59.718	20.685	59.260	20.397	0.457	0.288	0.008	Modified-IHS
Band 2	54.619	15.355	54.320	15.453	0.299	-0.099	0.005	
Band 3	60.908	12.558	60.629	13.011	0.280	-0.453	0.005	
AVERAGE	58.415	16.199	58.070	16.287	0.345	-0.088	0.006	
Band 1	59.718	20.685	59.260	20.397	0.457	0.288	1.457	Wavelet-IHS
Band 2	54.619	15.355	54.320	15.453	0.299	-0.099	1.299	
Band 3	60.908	12.558	60.629	13.011	0.280	-0.453	1.280	
AVERAGE	58.415	16.199	58.070	16.287	0.345	-0.088	1.345	

4.3. Comparison of Classification Results from Different Schemes

Classification was executed using RF and SVMs. Nine schemes were tested for image classification, and impact of variables was assessed. Two class labels were investigated: Buildings, and Non-Building. In the first step, Spot-5 and WorldView-2 are fused (i.e. using wavelet-PCA method) and the resulted image was used as a predictor variable for the classification process as the first scenario. In the second step, eight texture attributes were generated from the fused image using Envi software to form the GLCM image and the generated attributes (GLCM image) were used as predictor variable for the classification process as the second scenario. In the third step, nDSM is layer stacked with the fused image to form a new fused image with nDSM that was used as predictor variable for the classification process as the third scenario. Finally, the classifications were compared to each other based on kappa index, overall accuracy and *F1* measure. All these steps were repeated, for each fusion technique, once using RF and another one for SVMs. Tables 4 and 5 show comparison of the train model results using RF classification and SVMs classification for different fusion methods, respectively. Figures 6 and 7 illustrate visual comparison of classification maps using RF and SVMs for different fusion methods, respectively. SVMs were insensitive to the distribution of underlying data [17]. SVMs were able

to address high dimensionality problems and limited training samples. Also, it was able to generalize well, even with small data [17]. Other benefits comprise that no prior knowledge of the underlying data is required. On the other side, SVMs have a number of limits [48] such as it requires good kernel function and there is a scarcity of transparency in the outcomes. Whereas, RF advantages were ease of use, it requires fewer parameters and no intervention [49–51]. The main barrier of RF was that a large quantity of trees could make it too slow.

Table 4. Comparison of schemes using RF classification

Schemes	Overall accuracy	Kappa	F1 measure	No. of trees	Max features	Fusion method
Fused image using wavelet-IHS	0.944	0.851	0.926	81	2	Wavelet-IHS
GLCM raster of fused image	0.977	0.938	0.969	81	17	
Fused image +nDSM	0.961	0.893	0.946	27	3	
Fused image using wavelet-PCA	0.954	0.876	0.938	81	1	Wavelet-PCA
GLCM raster of fused image	0.972	0.923	0.961	27	17	
Fused image +nDSM	0.967	0.913	0.957	27	2	
Fused image using modified-IHS	0.940	0.839	0.919	81	2	Modified-IHS
GLCM raster of fused image	0.977	0.938	0.969	81	17	
Fused image +nDSM	0.958	0.885	0.943	27	3	

Table 5. Comparison of schemes using SVMs classification

Schemes	Overall accuracy	Kappa	F1 measure	Best gamma	Best C	Fusion method
Fused image using wavelet-IHS	0.954	0.877	0.938	1.0	10,000.0	Wavelet-IHS
GLCM raster of fused image	0.968	0.911	0.955	0.25	10.0	
Fused image +nDSM	0.965	0.904	0.952	8.0	100.0	
Fused image using wavelet-PCA	0.965	0.906	0.953	0.5	10,000.0	Wavelet-PCA
GLCM raster of fused image	0.973	0.927	0.963	0.0625	10.0	
Fused image +nDSM	0.973	0.926	0.963	0.5	10,000.0	
Fused image using modified-IHS	0.944	0.849	0.924	8.0	1000.0	Modified-IHS
GLCM raster of fused image	0.967	0.911	0.955	0.25	10.0	
Fused image +nDSM	0.958	0.885	0.943	1.0	10,000.0	

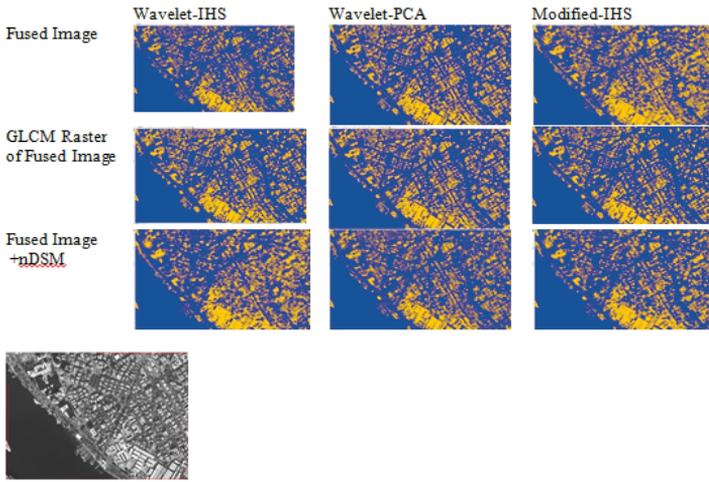


Fig. 6. Classification maps of different schemes for urban detection using random forest. The yellow colour represents buildings and the blue colour represents non-buildings

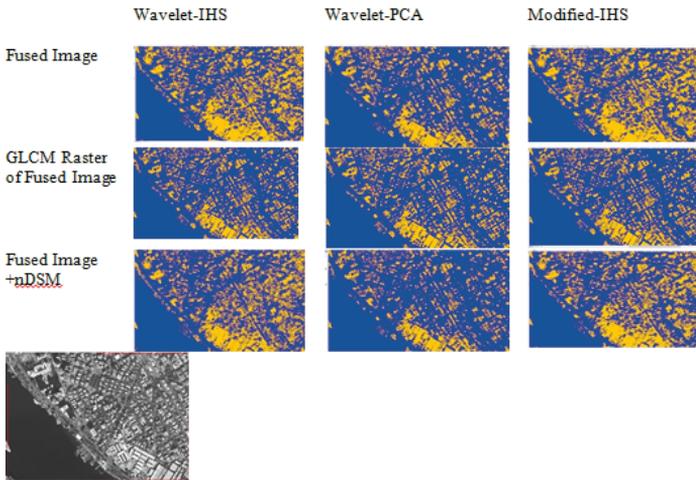


Fig. 7. Classification maps of different schemes for urban detection using SVMs. The yellow colour represents buildings and the blue colour represents non-buildings

Our results indicated RF might outperform SVM for image classification as in [52, 53].

It is clear that when textural features of colour images were used as input to the classifiers, the overall accuracy, *F1* measure and kappa are improved in most instances. This followed by integrating nDSM as a layer with the fused image comparing to using the fused image only. The results demonstrated that the difficulty in classifying spectral complex urban settings might be controlled by incorporating textural features.

Using fused images only overall accuracy (OA) reached 0.944, 0.954, 0.940 of wavelet-IHS, wavelet-PCA, and modified-IHS respectively using RF classification and OA reached 0.954, 0.965, 0.944 using SVMs classification.

Using textural features with fused images improved the accuracy of classification, and the OA reached 0.977, 0.972, 0.977 of wavelet-IHS, wavelet-PCA, and modified-IHS respectively using RF classification and OA reached 0.968, 0.973 and 0.967 using SVMs classification.

Using fused image plus nDSM revealed OA of 0.961, 0.967, 0.958 of wavelet-IHS, wavelet-PCA, and modified-IHS respectively using RF classification and OA reached 0.965, 0.973 and 0.958 using SVMs classification which is also improved the classification compared to using fused image only.

Based on the visual evaluation of the results we can state that almost all buildings were delineated successfully if they were larger than the expected minimum buildings area (30 and 50 m²).

In these experiments, the error was mainly caused by the merchant ships because the classification of ships depends on size and area, a factor which may be near or close to that of the building. Based on the area and size, the classifier mistakenly categorizes the ships as buildings.

5. Conclusion

The contribution of this research can be summarized as developing machine learning-based image fusion framework for building detection with its evaluation criteria. This model solved the building detection problem connected to the irregular structure and closeness of different buildings in urban areas. Three image fusion techniques were used to fuse Spot-5 and WorldView-2 images. The three fused images gave better interpretation than multispectral images. It is concluded that wavelet-PCA gave the least pixel errors and the three fused methods preserved the maximum sensible to the spectral quality of the original image.

To sum up, first, the impact of data labelling approach on the classification results. Second, our outcomes outperformed other peer methods. Third, a model to integrate spectral, height and textural information was proposed. It was found that textural features enhanced the accuracy of classification, and the overall accuracy reached 0.977, 0.972, 0.977 of wavelet-IHS, wavelet-PCA and modified-IHS respectively using random forest classification and overall accuracy reached 0.968, 0.973 and 0.967 using Support vector machines classification. Fourth, based on the results, random forest gave better results compared to support vector machines. Fifth, the results obtained in this study indicated that the developed procedure could be effectively used to delineate the boundaries of the rectangular shaped buildings, buildings with irregular shapes and quite small size ones with reliable accuracy. Sixth, it is recommended to use classifier ensemble for buildings extraction. Finally, the

proposed method is a valuable and effective for the prevention and trace of unorganized urbanization and can be used with other images.

Future work includes applying the developed model with the usage of other images in different locations.

References

- [1] Khalifa M.A.: *Redefining slums in Egypt. Unplanned versus unsafe areas*. Habitat International, vol. 35, 2011, pp. 40–49. <https://doi.org/10.1016/j.habitatint.2010.03.004>.
- [2] United Nations: *Sustainable Development Goals*. <https://www.un.org/sustainabledevelopment/cities/> [access: 7.06.2021].
- [3] San D.K., Turker M.: *Building extraction from high resolution satellite images using Hough transform*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, vol. XXXVIII, part 8, 2010, pp. 10–63.
- [4] Huang X., Zhang L.: *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, vol. 5, no. 1, 2012, pp. 161–172. <https://doi.org/10.1109/JSTARS.2011.2168195>.
- [5] Chaudhuri D., Kushwaha N.K., Samal A., Agarwal R.C.: *Automatic Building Detection from High-Resolution Satellite Images Based on Morphology and Internal Gray Variance*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, 2016, pp. 1767–1779. <https://doi.org/10.1109/JSTARS.2015.2425655>.
- [6] Uzar M.: *Automatic Building Extraction with Multi-sensor Data Using Rule-based Classification*. European Journal of Remote Sensing, vol. 47, 2014, pp. 1–18. <https://doi.org/10.5721/EuJRS20144701>.
- [7] Mishra A., Pandey A., Baghel A.S.: *Building Detection and Extraction Techniques: A Review*. [in:] *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 3816–3821.
- [8] Wang J., Qin Q., Chen L., Ye X., Qin X., Wang J., Chen C.: *Automatic Building Extraction from Very High-Resolution Satellite Imagery Using Line Segment Detector*. [in:] *2013 IEEE International Geoscience and Remote Sensing Symposium – IGARSS*, 2013, pp. 212–215. <https://doi.org/10.1109/IGARSS.2013.6721129>.
- [9] Wang M., Yuan S., Pan J.: *Building Detection in High-Resolution Satellite Urban Image Using Segmentation, Corner Detection Combined with Adaptive Windowed Hough Transform*. [in:] *2013 IEEE International Geoscience and Remote Sensing Symposium – IGARSS*, 2013, pp. 508–511. <https://doi.org/10.1109/IGARSS.2013.6721204>.
- [10] Liasis G., Stavrou S.: *Building Extraction in Satellite Images Using Active Contours and Colour Features*. International Journal of Remote Sensing, vol. 37, 2016, pp. 1127–1153. <https://doi.org/10.1080/01431161.2016.1148283>.

- [11] Sun G., Huang H., Weng Q., Zhang A., Jia X., Ren J., Sun L., Chen X.: *Combinational shadow index for building shadow extraction in urban areas from Sentinel-2A MSI imagery*. International Journal of Applied Earth Observation and Geoinformation, vol. 78, 2019, pp. 53–65. <https://doi.org/10.1016/j.jag.2019.01.012>.
- [12] Huang J., Zhang X., Xin Q., Sun Y., Zhang P.: *Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 151, 2019, pp. 91–105. <https://doi.org/10.1016/j.isprsjprs.2019.02.019>.
- [13] Wang X., Li P.: *Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 159, 2020, pp. 322–336. <https://doi.org/10.1016/j.isprsjprs.2019.11.028>.
- [14] Lu C., Yang X., Wang Z., Li Z.: *Using multi-level fusion of local features for land-use scene classification with high spatial resolution images in urban coastal zones*. International Journal of Applied Earth Observation and Geoinformation, vol. 70, 2018, pp. 1–12. <https://doi.org/10.1016/j.jag.2018.03.010>.
- [15] Liu J., Li T., Xie P., Du S., Teng F., Yang X.: *Urban big data fusion based on deep learning: An overview*. Information Fusion, vol. 53, 2020, pp. 23–133. <https://doi.org/10.1016/j.inffus.2019.06.016>.
- [16] Sohn G., Dowman I.: *Extraction of buildings from high-resolution satellite data*. [in:] Baltasvias E., Gruen A., Van Gool L. (eds.), *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, A.A. Balkema Publishers, Lisse 2001, pp. 345–355.
- [17] Sheykhmousa M., Mahdianpari M., Ghamisi M., Homayoun S.: *Support Vector Machine vs. Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, 2020, pp. 6308–6325. <https://doi.org/10.1109/JSTARS.2020.3026724>.
- [18] Belgiu M., Drăguț L.: *Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 96, 2014, pp. 67–75. <https://doi.org/10.1016/j.isprsjprs.2014.07.002>.
- [19] Deng C., Wu C.: *The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 86, 2013, pp. 100–110. <https://doi.org/10.1016/j.isprsjprs.2013.09.010>.
- [20] Chehata N., Guo L., Mallet C.: *Airborne lidar feature selection for urban classification using random forests*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 39, 2009, pp. 207–212.

-
- [21] Niemeyer J., Rottensteiner F., Soergel U.: *Contextual classification of lidar data and building object detection in urban areas*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 87, 2014, pp. 152–165. <https://doi.org/10.1016/j.isprsjprs.2013.11.001>.
- [22] Corcoran J., Knight J., Gallant A.: *Influence of multi-source and multitemporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota*. Remote Sensing, vol. 5, 2013, pp. 3212–3238. <https://doi.org/10.3390/rs5073212>.
- [23] Gislason P.O., Benediktsson J.A., Sveinsson J.R.: *Random forests for land cover classification*. Pattern Recognition Letters, vol. 27(4), 2006, pp. 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>.
- [24] Cherkassky V., Ma Y.: *Practical selection of SVM parameters and noise estimation for SVM regression*. Neural Networks, vol. 17, no. 1, 2004, pp. 113–126. [https://doi.org/10.1016/S0893-6080\(03\)00169-2](https://doi.org/10.1016/S0893-6080(03)00169-2).
- [25] Mountrakis G., Im J., Ogole C.: *Support vector machines in remote sensing: A review*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247–259, 2011. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.
- [26] Vågen T.-G.: *Remote sensing of complex land use change trajectories a case study from the highlands of Madagascar*. Agriculture, Ecosystems and Environment, vol. 115, 2006, pp. 219–228. <http://doi.org/10.1016/j.agee.2006.01.007>.
- [27] Saini A., Pratibha: *A Review on Various Techniques of Image Fusion for Quality Improvement of Images*. International Journals of Advanced Research in Computer Science and Software Engineering, vol. 8(1), 2018.
- [28] Wang X., Bai S., Li Z., Sui Y., Tao J.: *The PAN and MS image fusion algorithm based on adaptive guided filtering and gradient information regulation*. Information Sciences, vol. 545, 2021, pp. 381–402. <https://doi.org/10.1016/j.ins.2020.09.006>.
- [29] Wang X., Wang Y., Zhou C., Yin L., Feng X.: *Urban forest monitoring based on multiple features at the single tree scale by UAV*. Urban Forestry & Urban Greening, vol. 58, 2021, 126958. <https://doi.org/10.1016/j.ufug.2020.126958>.
- [30] Zhang Y, Sidibé D., Morel O., Mériaudeau F.: *Deep multimodal fusion for semantic image segmentation: A survey*. Image and Vision Computing, vol. 105, 2021, 104042. <https://doi.org/10.1016/j.imavis.2020.104042>.
- [31] Liu X., Jiao L., Li L., Tang X., Guo Y.: *Deep multi-level fusion network for multi-source image pixel-wise classification*. Knowledge-Based Systems, vol. 221, 2021, 106921. <https://doi.org/10.1016/j.knosys.2021.106921>.
- [32] Rasti B., Ghamisi P.: *Remote sensing image classification using subspace sensor fusion*. Information Fusion, vol. 64, 2021, pp. 121–130. <https://doi.org/10.1016/j.inffus.2020.07.002>.
- [33] ERDAS Imagine: *Geospatial Modeling & Visualization*. <https://gmvc.cast.uark.edu/photogrammetry/> [access: 2.01.2021].

- [34] Cao H., Tao P., Li H., Shi J.: *Bundle adjustment of satellite images based on an equivalent geometric sensor model with digital elevation model*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 156, 2019, pp. 169–183. <https://doi.org/10.1016/j.isprsjprs.2019.08.011>.
- [35] Fonseca L., Namikawa L., Castejon E., Carvalho L., Pinho C., Pagamisse A.: *Image Fusion for Remote Sensing Applications*. [in:] Zheng Y. (ed.), *Image Fusion and Its Applications*, IntechOpen Limited, London 2011. <https://doi.org/10.5772/22899>.
- [36] Ehlers M., Klonus S., Åstrand P.J., Rosso P.: *Multi-sensor image fusion for pan-sharpening in remote sensing*. International Journal of Image and Data Fusion, vol. 1(1), 2010, pp. 25–45. <https://doi.org/10.1080/19479830903561985>.
- [37] Ma J., Yu J., Zhang J., Zhang Y., Bi Q., Wang G., Yang J., Long Y.: *Processing Practice of Remote Sensing Image Based on Spatial Modeler*. [in:] 2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering, pp. 1–5. <https://doi.org/10.1109/RSETE.2012.6260666>.
- [38] Riyahi R., Kleinn C., Fuchs H.: *Comparison of different image fusion techniques for individual tree crown identification using quickbird images*. ISPRS Archives, vol. XXXVIII-1-4-7/W5, 2009, pp. 1–4.
- [39] Møller-Jensen L.: *Classification of urban land cover based on expert systems, object models and texture*. Computers, Environment and Urban Systems, vol. 21, no. 3/4, 1997, pp. 291–302. [https://doi.org/10.1016/S0198-9715\(97\)01004-1](https://doi.org/10.1016/S0198-9715(97)01004-1).
- [40] Haralick R.M., Shanmuga K., Dinstein I.: *Textural features for image classification*. IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3(6), 1973, pp. 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>.
- [41] Liu Y., Chen X., Wang Z., Jane Z.W., Ward R.K., Wang X.: *Deep learning for pixel-level image fusion: Recent advances and future prospects*. Information Fusion, vol. 42, 2018, pp. 158–173. <https://doi.org/10.1016/j.inffus.2017.10.007>.
- [42] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E.: *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.
- [43] Salah M., Trinder J.C., Shaker A., Hamed M., Elsagheer A.: *Integrating multiple classifiers with fuzzy majority voting for improved land cover classification*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVIII, part 3A, 2010, pp. 7–12.
- [44] Tricht K.V., Gobin A., Gilliams S., Piccard I.: *Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium*. Remote Sensing, vol. 10(10), 2018, 1642. <https://doi.org/10.3390/rs10101642>.
- [45] Elshehaby A.R., Taha L.G.: *A new expert system module for building detection in urban areas using spectral information and LIDAR data*. Applied Geomatics, vol. 1(4), pp. 97–110, 2009. <https://doi.org/10.1007/s12518-009-0013-1>.

-
- [46] Maxwell A.E., Warner T.A., Guillén L.A.: *Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies – Part 1: Literature Review remote sensing*. Remote Sensing, vol. 13(13), 2021, 2450. <https://doi.org/10.3390/rs13132450>.
- [47] Zhang X., Han L., Han L., Zhu L.: *How Well Do Deep Learning-Based Methods for Land Cover Classification and Object Detection Perform on High Resolution Remote Sensing Imagery?* Remote Sensing, vol. 12(3), 2020, 417. <https://doi.org/10.3390/rs12030417>.
- [48] Mohamed A.E.: *Comparative Study of Four Supervised Machine Learning Techniques for Classification*. International Journal of Applied Science and Technology, vol. 7, no. 2, 2017, pp. 5–18.
- [49] Ming D., Zhou T., Wang M., Tan T.: *Land cover classification using random forest with genetic algorithm-based parameter optimization*. Journal of Applied Remote Sensing, vol. 10, no. 3, 2016, 035021. <https://doi.org/10.1117/1.JRS.10.035021>.
- [50] Jhonnerie R., Siregar V.P., Nababan B., Prasetyo L.B., Wouthuyzen S.: *Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries*. Procedia Environmental Sciences, vol. 24, 2015, pp. 215–221. <https://doi.org/10.1016/j.proenv.2015.03.028>.
- [51] Rao K.V.R., Kumar P.R.: *Land Cover Classification Using Sentinel-1 SAR Data*. International Journal for Research in Applied Science and Engineering Technology, vol. 5, no. 12, 2017, pp. 1054–1060.
- [52] Belgiu M., Drăguț L.: *Random forest in remote sensing: A review of applications and future directions*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, 2016, pp. 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- [53] Delgado F., Cernadas E., Barro S., Amorim D.: *Do we need hundreds of classifiers to solve real world classification problems?* Journal of Machine Learning Research, vol. 15, no. 1, 2014, pp. 3133–3181.