Journal of the Indian Society of Remote Sensing
https://doi.org/10.1007/s12524-020-01235-z

RESEARCH ARTICLE

Automatic Extraction of Buildings from UAV-Based Imagery Using Artificial Neural Networks

Prakash Pilinja Subrahmanya[1] · Bharath Haridas Aithal[1] · Satarupa Mitra[2]

Received: 6 October 2020 / Accepted: 22 October 2020

Abstract
The rapid growth of major cities across the counties demands accurate building extraction techniques. Unmanned Aerial Vehicles (UAV) help in obtaining terrain information that can be used to extract urban features. Recent advancement has led to the capture of aerial images of the earth surface in micro-detail using UAV. These aerial images enable us to perform classification, feature extraction, and height estimation at a finer scale. In this work, aerial images of the university campus were captured using a quadcopter drone equipped with high-resolution camera and satellite navigation system. Approximately 500 images were captured in the study area with necessary overlap and side lap. Captured images were subjected to aerial triangulation, dense image matching, and point cloud generation to produce Digital Surface Models (DSM) and orthophoto. Various machine learning algorithms—random forest (RF), support vector machine (SVM), naïve Bayes (NB) and artificial neural networks (ANN)—have been used to extract building rooftops from derivatives of UAV-captured imageries, and accuracies were compared. Algorithms were trained using both spectral and elevation information to extract building rooftops, and improvements shown due to the addition of elevation data in training the model are observed. The proposed method is aimed at improving building-level information extraction and providing accurate building information to aid authorities for better planning and management.

Keywords Artificial neural network · Building extraction · Machine learning · SVM · RF · UAV

Introduction

Urbanization is a complex and irreversible phenomenon that is basically due to the migration of people to cities in pursuit of employment, education facilities, and development of housing, industries, transportation infrastructure, etc. (Bharath et al. 2017). Urban areas constitute the majority of paved surfaces representing complex nature corresponding to varied surface types and heights in a remotely sensed image. Rapid redevelopment activities in major cities across the counties demand accurate building extraction techniques. Although space-based sensors are capable of proving high-resolution imagery, UAV (drone)-captured imagery gives the sophistication of cloud-free and timely data. Satellite-based position-enabled drones provide the advantage of generating accurate DSM from remotely sensed images. Drone-captured images are useful for surface feature extraction in various types of land use with the highest accuracy (Becker et al. 2017). Recent advancements have led to capture aerial imageries of the land in micro-detail using drones. These data-heavy aerial images enable us to perform classification, feature extraction, and height estimation at a finer scale (Jarzabek and Karpina 2016).

Previously several techniques have been developed to extract buildings from remotely sensed imageries, viz. clustering, a supervised classification, object-based classification, etc. Huang et al. 2017 used clustering techniques to classify various land use types and label them later according to land-use class. Mattupalli et al. 2018 performed supervised method of classification called maximum likelihood to delineate various land features. Object-

✉ Prakash Pilinja Subrahmanya
prakash.ps@iitkgp.ac.in; psprakashgis@gmail.com

[1] Ranbir and Chitra Gupta School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India

[2] Symbiosis Institute of Geoinformatics, Symbiosis Institute of Technology, Pune, Maharashtra 412115, India

Published online: 06 November 2020

based classification resulted in good outcomes from drone-captured imagery consisting of multispectral data (Kakooei and Baleghi 2017). Becker et al. 2017 and Kahraman 2017 employed machine learning techniques to do classification based on RGB information. In the project work, we carried out building extraction by various machine learning methods (viz. random forest (RF), support vector machine (SVM), naïve Bayes (NB) and artificial neural networks (ANN)) and results were compared with the ground truth map. The methods adopted have several advantages such as requires minimal training samples, computationally fast, and well developed.

Data Utilized

Aerial images of the university campus were captured using a quadcopter drone equipped with a 20 MP CMOS camera and GLONASS/GPS satellite navigation system. A total of 530 images of dimensions 5280 X 3956 were captured in the study area such that forward overlap is not less than 75% and side overlap is about 65%. The overlapping of the aerial images helps to perform the photogrammetric process to extract a DSM. Imageries are acquired with a ground sample distance of 3.8 cm, across three bands: blue, green, and red having 8 bits data radiometry. Figure 1 shows four such images captured using the drone as examples. A subset of the entire campus measuring an area of 24.8 acres is clipped from generated orthophoto and DSM to perform classification.

Machine Learning Techniques

Paradigm shifts in image processing took place with the development of machine learning techniques and availability of high computing capacities. Various machine learning algorithms have proven to be efficient in classification and feature extraction from satellite (Thanh and Kappas, 2018) and aerial-based images (Piragnolo et al. 2017). All the four techniques described below are supervised learning methods in which the model is trained and class prediction is performed.

Random Forest

RF is an ensemble of decision tree classifier that gives multiple outputs, and the result is predicted based on the statistical mode of the resultants (Breiman 2001). Decision tree classifier works based on a series of test questions and conditions arranged in a tree structure, and they are binary. Random forests are capable of classifying remote sensing imagery because of its strength in dealing with an outlier in training data (Horning 2010). Mathematical formulations of the random forest type of classification are explained in Das et al. 2018.

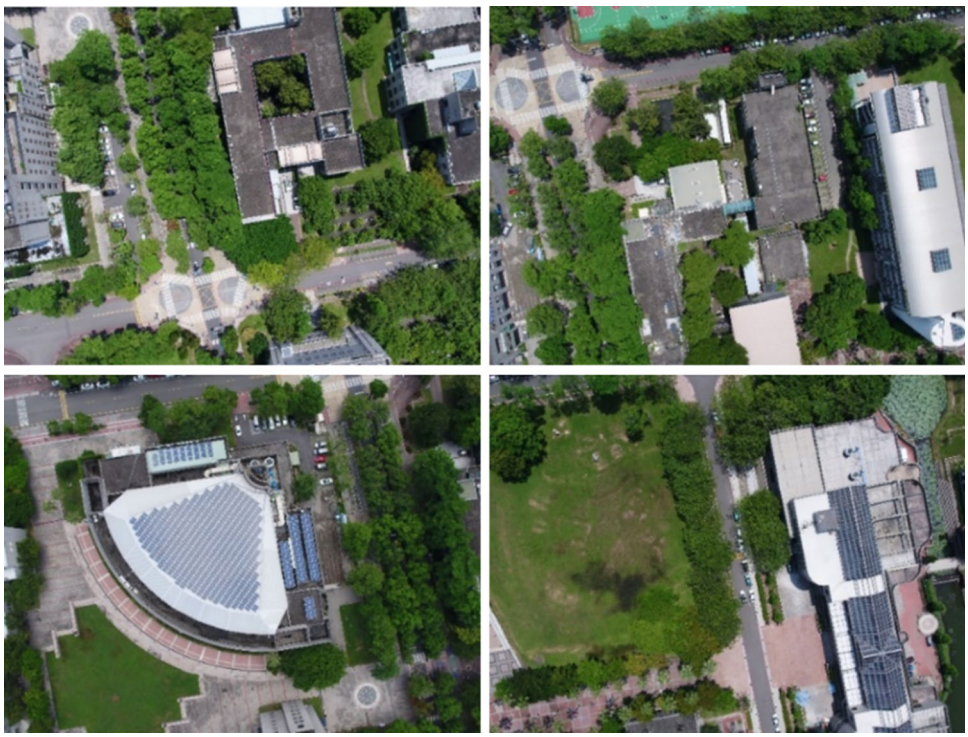

Fig. 1 Drone-captured images of the university campus

Support Vector Machine

Cortes and Vapnik 1995 developed this method based on the concept of maximizing the margins between the hyperplane and closest data points. SVM has been used in the classification of remote sensing data in several earlier works (Dixon and Candade 2008; Benarchid et al 2013; Prakash et al. 2018). The mathematical formulation of the SVM method is given in the literature (Foody and Mathur 2004; Prakash 2018), and SVM classifier gives the flexibility of using four variety of kernels, namely sigmoid, radial basis function, linear and polynomial. Among nonlinear methods, feature vectors are projected into higher dimensions of space to classify the data efficiently. In the current project, we utilized the radial basis function as it is capable of separating nonlinear data.

Naïve Bayes

Naïve Bayes algorithm is based on Bayes theorem, i.e., the probability of occurrence of event A given that event B occurs, as expressed in the equation given below. Here assumption made is that A and B are independent of each other. The mathematical formulation of the classifier and the use case were given by Aksoy et al. 2005. In a paper, Lv et al. 2017 showed the preparation of land use maps using naïve Bayes classifier with high-resolution aerial imagery.

Artificial Neural Network

A multilayer perceptron (MLP) as ANN is used for classification of the orthophoto to extract buildings. MLP model in its basic form consists of three layers, namely an input layer, a hidden layer, and an output layer, and each layer is connected to the next layer either fully or randomly (Srivastava et al. 2012; Mokhtarzade and Zoej 2007). In MLP, the sum of input values x_i is passed onto each neuron j in the hidden layer weighted by corresponding weight w_{ij} and generated its output y_i as function of addition.

$$y_i = f\left(\sum W_{ji} X_i\right)$$

In the above equation, f is called activation function, it can be a step function or radial basis or sigmoidal function. The difference between desired and actual values of the output neuron e is given by

$$e = \sum \left(y_{dj} - y_j\right)^2$$

y_{dj}: desired value of output, y_j: actual value of output.

The weight W_{ij} is adjusted at every iteration to minimize the value of e using back propagation techniques. Neurons of input layers hold various input features such as training class or bands of imagery data, and neurons of the output layer represent land use classes in remote sensing image classification. In the current work, a single layer back-propagation neural network (Somers and Casal 2009) was used to classify imagery into binary classes.

Method

The overall method formulated for the study is shown in Fig. 2; it includes four steps. Firstly, aerial images were captured using a camera fitted to a drone, then image data (orthophoto and DSM) was prepared for the entire campus later, and machine learning algorithm was applied to the data to perform segmentation. The output of the segmentation was subjected to accuracy assessment with the ground truth map.

Data Capture

In the data capture stage, flight planning parameters were adjusted by mentioning ground sample distance, forward overlap, and side overlap, and the boundary was fixed according to the chosen area that we were required to record. The flight plan tool helped us to visualize the estimated parameters to take precautionary measures and fix technical properties of the image to be captured, viz. flying height, time of flight, and number of images. Following the successful flying, data were transferred to a computing machine for further processing.

Data Preparation

A point cloud model is prepared from drone-captured imageries by aligning them based on camera position and orientation. The DSM is generated from a dense point cloud that represents height information as regular grids. The surface model and aligned photographs are utilized in preparing orthophoto with a spatial resolution of 0.3 cm, containing RGB bands. Training areas representing building and non-building categories are manually digitized as a separate feature class, and pixels corresponding to training areas are extracted from orthophoto and DSM to train the machine learning models mentioned in Sect. 3.

Classification

Supervised classification is carried out using various machine learning methods, viz. random forest, naïve Bayes, support vector machine, and artificial neural network. All machine learning algorithms were trained with the manually labeled dataset digitized from the orthophoto; Fig. 3 shows the generated orthophoto and DSM. During

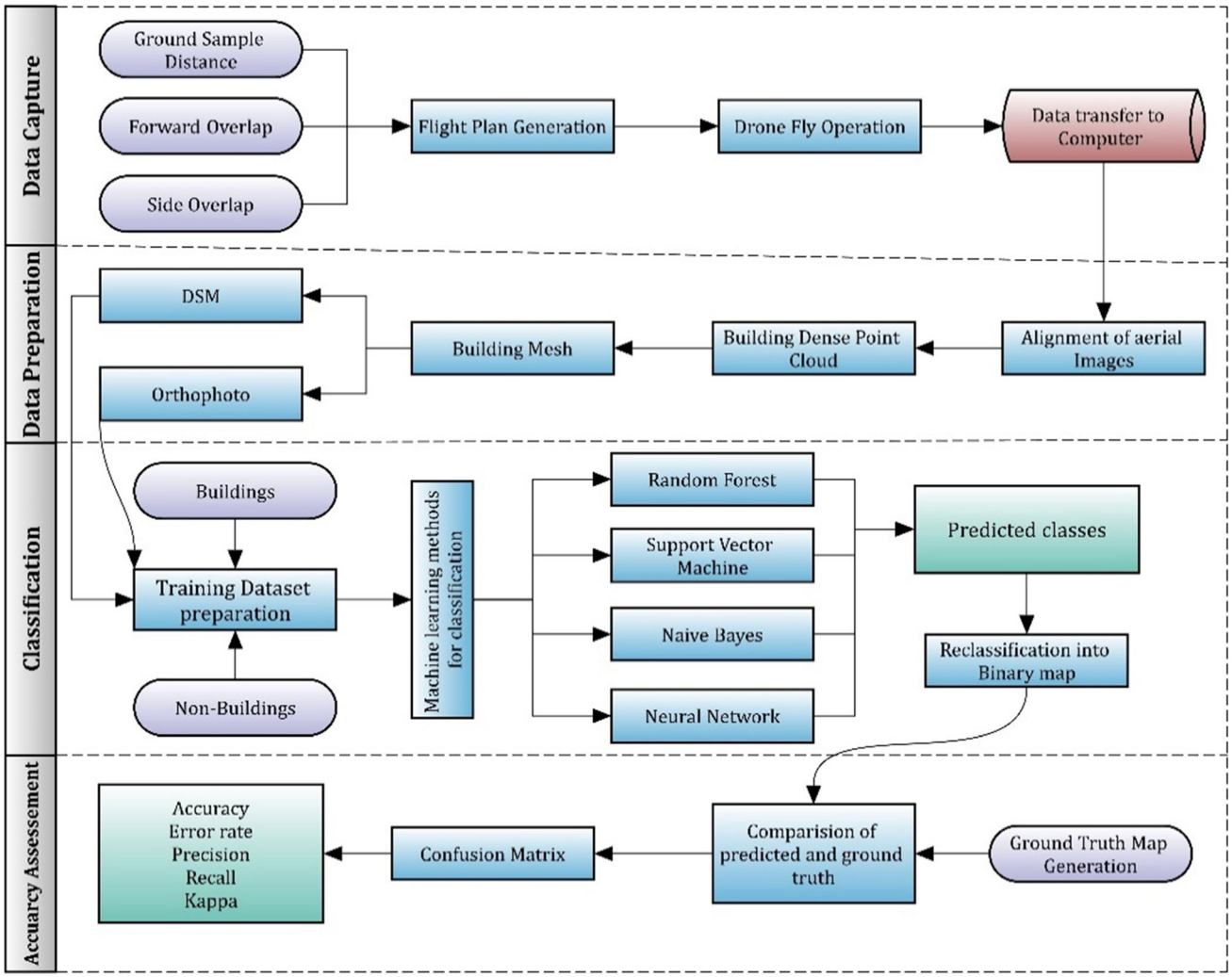

Fig. 2 Method adopted for the study

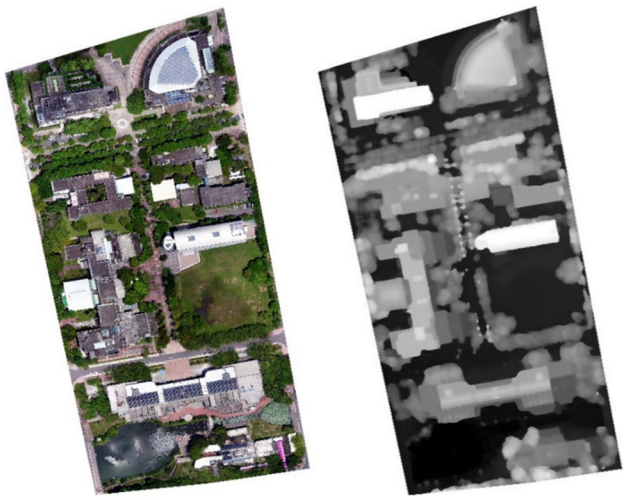

Fig. 3 Orthophoto (left) and digital surface model (right)

training, we made use of ancillary data (DSM) in addition to image data (RGB), which helps to increase the classification accuracy. A binary classification process was carried out with four algorithms using the prepared set of training data. The predicted output from each algorithm was reclassified into two classes, namely buildings and non-buildings, by manually identifying a threshold value from the classified output. It is observed that SVM took a longer time than the other algorithms in fitting the model as well as for prediction and the Bayesian method performed fastest among all for fitting as well as prediction.

Accuracy Assessment

A confusion matrix was generated (Table 1) for classified maps concerning ground truth map that was prepared by manual digitization for the entire study area. Specificity (true negative rate) is the measure of how often does predictor predict non-buildings when is non-building. Recall

Table1 The confusion matrix used for accuracy evaluation

		Classifier prediction	
		Buildings	Non-buildings
Actual class	Buildings	True positive (TP)	False negative (FN)
	Non-buildings	False positive (FP)	True negative (TN)

TP Building pixels predicted as buildings, *TN* Non-building pixels predicted as non-buildings, *FP* Non-building pixels predicted as buildings, *FN* Building pixels predicted as non-buildings

(true positive rate) is how often the predictor predicts buildings as buildings. Error rate indicates the overall misclassification rate, and accuracy is the measure of correct classification. Kappa is the measure of agreement of accuracy (Rwanga and Ndambuki 2017). The performance metric for segmentation tasks called Intersection over Union (IoU) is calculated to the extracted building area.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Error rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Kappa} = \frac{N*d - Q}{N^2 - Q}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

Here, N = Total number of samples (TP + TN + FP + FN), d = Sum of diagonal elements (TP + TN), Q = X*Y, X = Row sum for buildings (TP + FN), Y = Row sum for non-buildings (FP + TN).

Results and Discussion

An orthophoto is generated from drone-captured imageries consisting of three bands: blue, green, and red, and resolution or pixel size is 0.38 cm. The dimensions of the orthophoto are 22,373 X 22,120; the assigned coordinate reference system is WGS 84 and occupies a disc space of 453 megabytes. The DSM generated using dense point cloud is a single-layer raster representing height variations from 94 to 139 m; Fig. 3 shows both orthophoto and DSM. Statistical measures of orthophoto and DSM are given in Table 2. Classified binary output and extracted building area were superimposed on orthophoto as shown in Fig. 4. A vectorization process was performed, during which a small group of pixels and outliers were filtered out. We could clearly see that the corners of a few buildings were not captured correctly due to tree occlusion.

Table 2 Descriptive measures of data used for classification

	Blue	Green	Red	DSM
Maximum	255	255	255	139.42
Minimum	0	0	0	94.03
Mean	111.52	119.42	99.03	115.33
Standard deviation	59.38	52.33	66.97	7.02

The predicted map consists of values from −1 to + 1 (Fig. 4), and careful observation was required to identify the threshold value. In the current study, we tested for several values 0.2 to 0.9 on a trial and error basis among various outputs and the best performing threshold was identified. Values above the threshold were assigned as buildings, and values below the threshold were assigned as non-buildings. It was found that a single-layer neural network performed efficiently than other methods in all measures with a minimum error rate of 0.07. Improvement in overall accuracies due to the addition of ancillary data was 13.2, 15.8, 12, and 11.3 percentages in RF, NB, SVM, and NN, respectively. Accuracies of various algorithms are tabulated in Table 3. Results show the increase in the overall accuracy of various methods due to supplementing training with ancillary data like DSM. Also, all other measures of accuracies were improved with additional training, and overall improvement was found in the order of NB, RF, SVM, and NN, respectively.

Conclusion

The study demonstrates the potential of UAV-captured imageries to extract buildings using various machine learning algorithms. The current work employed the four most popular machine learning methods to perform binary classification of an orthophoto, hence to delineate building features. The developed models were found to be capable of extracting buildings in the scene but failed at places where occlusion happened due to tree cover.

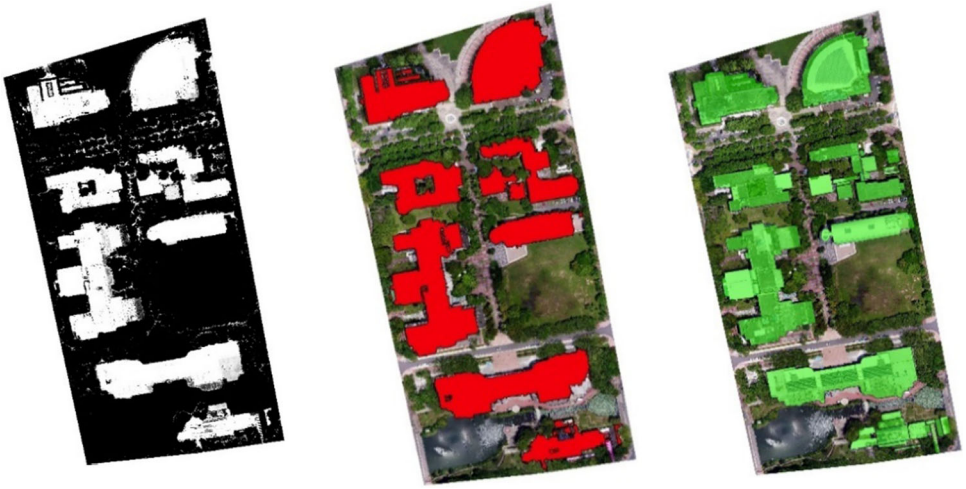

Fig. 4 Buildings class predicted as extracted by ANN (left), buildings roofs after applying a filter(center), building digitized or ground truth(right)

Table 3 Accuracy metrics of various machine learning algorithms used in the study

Algorithm	RF		NB		SVM		NN	
Data used	RGB	RGB& DSM	RGB	RGB& DSM	RGB	RGB& DSM	RGB	RGB& DSM
Accuracy	0.79	0.91	0.74	0.88	0.81	0.92	0.82	0.93
Error rate	0.21	0.09	0.26	0.12	0.19	0.08	0.18	0.07
Recall	0.83	0.94	0.80	0.91	0.84	0.97	0.88	0.96
Specificity	0.88	0.84	0.58	0.81	0.73	0.83	0.70	0.87
Kappa	0.74	0.89	0.68	0.85	0.77	0.90	0.78	0.91
IoU	0.75	0.88	0.70	0.84	0.77	0.89	0.78	0.90

Misclassification occurred in some open features such as pavement, concrete and parking lots, which were classified wrongly as buildings by the algorithms. The use of ancillary data has reduced misclassification, and hence overall error rate by 38%, and accuracy was increased by 11.3% in the neural network method. The single-layer backpropagation neural network, additionally trained with ancillary data, achieved classification accuracy up to 93%. Further improvement in the algorithm is required to extract buildings shape accurately and to overcome problems of occlusion by trees and misclassification at places. The proposed method is aimed at improving building-level information extraction and providing accurate building information for authorities for better planning and management.

Acknowledgements We are thankful to SERB, India, The Ministry of Science and Technology, Government of India, Ranbir and Chitra Gupta School of Infrastructure Design and Management and Sponsored research in Consultancy cell (SRIC), Indian Institute of Technology, Kharagpur, and Department of Science and Technology, West Bengal, for the financial and infrastructure support.

References

Aksoy, S., Koperski, K., Tusk, C., Marchisio, G., & Tilton, J. C. (2005). Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing, 43*(3), 581–589.

Becker, C., Häni, N., Rosinskaya, E., d'Angelo, E., & Strecha, C. (2017). Classification of aerial photogrammetric 3D point clouds. arXiv preprint arXiv:1705.08374.

Benarchid, O., Raissouni, N., El Adib, S., Abbous, A., Azyat, A., Achhab, N. B., & Chahboun, A. (2013). Building extraction using object-based classification and shadow information in very high-resolution multispectral images, a case study: Tetuan, Morocco. *Canadian Journal on Image Processing and Computer Vision 4*(1), 1–8.

Bharath, H. A., Chandan, M. C., Vinay, S., & Ramachandra, T. V. (2017). Modelling urban dynamics in rapidly urbanising Indian cities. *Egyptian Journal of Remote Sensing and Space Science 21*(3), 201–210.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning 20*(3), 273–297.

Das, S. K., Bharath, H. A., & Prakash, P. S. (2018). Automated building extraction using high-resolution satellite imagery though Ensemble modelling and Machine learning. *Remote Sensing of Land*, *2*, 1–14.

Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: one or the

other, or both? *International Journal of Remote Sensing, 29*(4), 1185–1206.

Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, *93*(1–2), 107–117.

Horning, N. (2010, December). Random Forests: An algorithm for image classification and generation of continuous fields data sets. In: *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, Osaka, Japan, (Vol 911).

Huang, L., Yu, X., & Zuo, X. (2017). Edge detection in UAV remote sensing images using the method integrating zernike moments with clustering algorithms. *International Journal of Aerospace Engineering*, *2017*, 1793212. https://doi.org/10.1155/2017/1793212.

Jarzabek-Rychard, M., & Karpina, M. (2016). Quality analysis on 3D building models reconstructed from UAV imagery. International Archives of the Photogrammetry. *Remote Sensing & Spatial Information Sciences*, *41*, 1121–1126.

Kahraman, I. (2017). An approach for road network detection from satellite images using neural networks. *International Journal of Soft Computing and Artificial Intelligence, 1,* 15–18.

Kakooei, M., & Baleghi, Y. (2017). Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment. *International journal of remote sensing, 38*(8–10), 2511–2534.

Lv, Z., Zhang, P., & Atli Benediktsson, J. (2017). Automatic object-oriented, spectral-spatial feature extraction driven by tobler's first law of geography for very high-resolution aerial imagery classification. *Remote Sensing, 9*(3), 285.

Mattupalli, C., Moffet, C., Shah, K., & Young, C. (2018). Supervised classification of RGB aerial imagery to evaluate the impact of a root rot disease. *Remote Sensing, 10*(6), 917.

Mokhtarzade, M., & Zoej, M. V. (2007). Road detection from high-resolution satellite images using artificial neural networks. *International journal of applied earth observation and geoinformation, 9*(1), 32–40.

Piragnolo, M., Masiero, A., & Pirotti, F. (2017). Open source R for applying machine learning to RPAS remote sensing images. *Open Geospatial Data, Software and Standards, 2*(1), 16.

Prakash, P. S., Soumya, K. D., & Bharath, H. A. (2018, November). Urban building extraction using satellite imagery through Machine Learning. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1670–1675), IEEE.

Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. *International Journal of Geosciences, 8*(04), 611.

Somers, M. J., & Casal, J. C. (2009). Using artificial neural networks to model nonlinearity: The case of the job satisfaction—job performance relationship. *Organizational Research Methods, 12*(3), 403–417.

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research, 50*(9), 1250–1265.

Thanh Noi, P., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors, 18*(1), 1.