Assessing the contribution of RGB VIs in improving building extraction from RGB-UAV images

Richmond Akwasi Nsiah *[1], **Saviour Mantey** [1], **Yao Yevenyo Ziggah** [1]

[1] University of Mines and Technology, Department of Geomatics Engineering, Tarkwa, Ghana, ransiah94@gmail.com, smantey@umat.edu.gh, yyziggah@umat.edu.gh

Abstract

Buildings are a fundamental component of the built environment, and accurate information regarding their size, location, and distribution is vital for various purposes. The ever-increasing capabilities of unmanned aerial vehicles (UAVs) have sparked an interest in exploring various techniques to delineate buildings from the very high-resolution images obtained from UAV photogrammetry. However, the limited spectral information in UAV images, particularly the number of bands, can hinder the differentiation between various materials and objects. This setback can affect the ability to distinguish between different materials and objects. To address this limitation, vegetative ındices (VIs) have been employed to enhance the spectral strength of UAV orthophotos, thereby improving building classification. The objective of this study is to evaluate the contribution of four specific VIs: the green leaf index (GLI), red-green-blue vegetation index (RGBVI), visual atmospherically resistant index (VARI), and triangular greenness index (TGI). The significance of this contribution lies in assessing the potential of each VI to enhance building classification. The approach utilized the geographic object-based image analysis (GeoBIA) approach and a random forest classifier. To achieve this aim, five datasets were created, with each dataset comprising the RGB-UAV image and a corresponding RGB VI. The experimental results on the test dataset and a post-classification assessment indicated a general improvement in the classification when the VIs were added to the RGB orthophoto.

1. Introduction

Among the myriad of urban features, buildings represent a fundamental component [1]. As such, obtaining accurate and detailed information on buildings is crucial for urban planning, infrastructure development, disaster management, and other applications [2]. In remote sensing, the term 'building extraction' is used to describe the process of delineating building footprints or roof outlines from remotely sensed data, such as very high-resolution (VHR) aerial and satellite images. These datasets are the most widely used as they offer rich spatial details. Furthermore, the enhanced spatial resolution of these datasets improves the ability to distinguish various objects in urban settings, thereby facilitating the extraction of individual building information [3].

The recent advancements in unmanned aerial vehicle (UAV) technologies, coupled with the sophistication of imaging sensor systems, have resulted in an increased use of these systems to capture aerial images of areas of interest. The VHR images obtained can be quickly processed to obtain orthophotos, thus providing an alternative dataset to extract building outlines [4]. This development has sparked an interest in exploring various methods to delineate building objects from VHR UAV imagery.

Conventional methods for building extraction involve manual approaches, which include delineating the outlines of buildings using various computer-aid designs (CAD). Also, rule-based techniques that leverage knowledge of buildings have been employed in building extraction. While the manual approach shows promise, it is repetitive and time-consuming when applied to larger areas [5-6]. In contrast, rule-based techniques rely on straight lines and use low-level edge detection and perceptual grouping to extract building outlines [7-8]. However, these line-based approaches encounter limitations with certain building geometries and struggle with low signal-to-noise ratio (SNR) in VHR images [9–10]. Some methods employ template matching, which involves the use of manually generated templates and similarity measures for building extraction [11–14]. Despite their flexibility and accuracy, template-based approaches require extensive prior knowledge of the geometrical shape parameters for the design and

generation of the templates. Moreover, the need to develop different templates for different applications increases the computational costs and reduces the ability to process extensive data [9, 11].

Other studies utilise knowledge of building geometries, such as box-shape, T-shape, L-shape, or E-shape, and the spatial or contextual relationship between buildings and the background, such as shadows, are used to delineate buildings [15–19]. The drawback of these approaches is their reliance on building knowledge, and the task of transferring implicit knowledge into explicit detection rules is challenging. If the rules are too strict, buildings may be missed, whereas overly loose rules can lead to many false detections [9–11]. Furthermore, several factors can make the extraction of buildings using rule-based techniques a challenging task. These factors include such as the intricate shape and size of buildings, occlusion, and imaging angles.

Contrastingly, image classification approaches (pixel-based and object-based) categorize pixels in images into specific classes, primarily buildings, and non-buildings in the building extraction domain [20]. Pixel-based classification operates by examining pixels in isolation and leverages the spectral characteristics of each pixel to assign them to distinct classes. While this approach is relatively simple, it encounters difficulties with spatial variability when applied to VHR images. On the other hand, the object-based approach groups image pixels into spectrally homogenous segments using various image-segmentation algorithms. Subsequently, an algorithm is adopted to classify the segments into predefined categories. This transition from focusing on individual pixels to evaluating at the segment level introduces complexity into the classification process. As a result, this approach can capture spatial relationships and complex patterns in VHR images [21–22]. The object-based approach, alternatively referred to as geographic object-based image analysis (GeoBIA), has emerged as an efficient approach for automating the extraction of objects from remote sensing data [23]. GeoBIA integrates image segmentation algorithms to segment VHR images into image objects, extract and employs machine learning algorithms such as support vector machine (SVM), random forest (RF), and decision trees (DT), among others, to classify the image objects, based on their spatial information, and spectral characteristics. This process makes GeoBIA particularly suited for building extraction [24 – 25].

Several researchers have since explored the use of GeoBIA for extracting buildings from VHR images. Aminipouri et al. [24] conducted a study that leveraged VHR satellite imagery to extract spatial information about slum settlements using object-oriented techniques. The study aimed to determine the feasibility of using VHR orthophotos to create an accurate inventory of buildings for estimating the slum population. The researchers used eCognition software for image segmentation and classification of building roofs in three different slum areas in Dar-Es-Salaam. The proposed model achieved a roof extraction accuracy of 91%. The estimated population represented 82.2%, 72.5%, and 68.3% for the wards of Charambe, Manzese, and Tandale, respectively. In another study, Benarchid et al. [26]

presented an automatic building extraction approach in Tetuan City. This method employed an object-based classification and shadow information derived from VRH multispectral images. The shadow information was extracted using invariant color features. The quality assessment was performed at two different levels: area and object. The area level evaluated the building delineation performance, whereas the object level assessed the accuracy in the spatial location of individual buildings. The results showed an overall building detection percentage of 87.60% when the parameters were properly adjusted and adapted to the type of areas considered.

The methodology adopted by Chen et al. [9] presented an object-based and machine learning-based approach for automatic house detection from RGB high-resolution images. The study utilised thresholding, watershed transformation, and hierarchical merging for image segmentation. In addition, the study proposed two new features, namely edge regularity indices (ERI) and shadow line indices (SLI), to capture the characteristics of house regions effectively. The researchers employed three classifiers, namely AdaBoost, random forests, and support vector machine (SVM), to identify houses from test images. The proposed ERI and SLI features improved the precision and recall by 5.6% and 11.2%, respectively. Norman et al. [22] focused on urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. The study employed a medium-resolution Sentinel-2B image and applied SVM and DT algorithms for the classification of buildings. The study underscored the significance of segmentation parameters and feature selection, with SVM outperforming DT and achieving an accuracy of 93%.

Frishila and Kamal [27] aimed to examine the effectiveness of spectral features in identifying and mapping building objects and assess the accuracy of the mapping result. The location of the study sample was in parts of Padang City, West Sumatra, and the image used was a pan-sharpened GeoEye-1 image (0.5 m pixel size). Image segmentation was done by the multi-resolution segmentation method to delineate candidate segments for building objects. Each segment was then assigned to building and non-building classes by applying a rule-based classification algorithm. Several spectral features were incorporated in discriminating the objects, including several band ratios that involve all bands in GeoEye-1 (Blue, Green, Red, and near-IR), iron oxide indices, mean value of red and NIR bands, border contrast of red and NIR bands, HIS, and Quantile of the bands. The map result indicates that building and non-building objects could be separated using spectral features of the GeoEye-1 image. However, there are some classification inaccuracies, mainly for the densely populated urban areas where buildings objects are close to each other. An area-based accuracy assessment indicated that the use of spectral features provides an overall accuracy of 68.7%.

Hossain and Chen [28] introduced several modifications to previously proposed hybrid segmentation methods for building extraction. They used the reference polygon to identify optimal parameters, a donut-filling technique to reduce over-segmentation

caused by roof elements, and illumination differences to restrict merging with shadow. Their methodology was tested on a UAV image with visible bands only and achieved better results compared to other methods. One of the strengths of their method was that there was no parameter tuning and user interaction at running time. In addition, it was able to segment both small and large buildings without using any scale or object size parameters.

Dornaika et al. [29] presented a generic framework that exploits recent advances in image segmentation and region descriptor extraction for the automatic and accurate detection of buildings on aerial orthophotos. The proposed solution is supervised in the sense that the appearances of buildings are learnt from examples. For the first time in the context of building detection, they used the matrix covariance descriptor, which proved to be very informative and compact. They introduced a principled evaluation that allows selecting the best pair segmentation algorithm-region descriptor for the task of building detection. The proposed approach presents several advantages in terms of scalability, suitability, and simplicity with respect to the existing methods. Furthermore, the proposed scheme (detection chain and evaluation) can be deployed for detecting multiple object categories that are present in images and can be used by intelligent systems requiring scene perception and parsing, such as intelligent unmanned aerial vehicle navigation and automatic 3D city modeling.

Argyridis and Argialas [30] developed a GEOBIA approach that integrates Deep Learning classification and Fuzzy Ontologies to monitor building changes in suburban areas of Greece. They employed deep belief networks (BDNN) on the lowest level of the segmentation hierarchy for the initial detection of areas of possible change. The classification result was then refined based on interpretation rules developed on the upper levels of the hierarchy. Their accuracy assessment indicated that 93.5% of the total number of changes were successfully detected, while the commission error was less than 20%.

While GeoBIA has shown considerable promise in building classification and segmentation, the spectral limitations of UAV-RGB imagery pose a challenge, especially when distinguishing between buildings and other urban features (3). Researchers have since used various ancillary datasets to address this drawback when VHR multispectral images are adopted for building extraction, the near-infrared (NIR) information has been established to be highly effective in differentiating vegetation from buildings [4]. Most often, various spectral indices such as the normalised difference vegetation index (NDVI), normalised difference built-up index (NDBI), and the normalised difference water index (NDWI), among others, are utilised to enhance building-background separability, thereby improving building extraction [31-33].

To address this challenge, many researchers have resorted to ancillary datasets, such as the RGB vegetative indices (VIs), to distinguish buildings from vegetative features [1, 4]. VIs can capture subtle spectral variations, and they present a promising approach for enhancing building classification and segmentation when integrated with the spatial context analysis of GeoBIA [1, 34].

While some research works have concentrated on improving building classification using RGB VIs, a comprehensive comparison evaluating the impact of each VI on classification accuracy has yet to be conducted. This study aims to bridge this gap. Consequently, the primary objective of this study is to investigate the impacts of incorporating RGB-based VIs into the GeoBIA classification pipeline for building extraction. To achieve this objective, four well-established VIs were utilized. These include the green leaf index (GLI), red-green-blue vegetation index (RGBVI), visual atmospherically resistant index (VARI), and triangular greenness index (TGI) were employed. These VIs were combined with UAV-RGB imagery to form separate datasets for the building classification task. The efficacy of each dataset was evaluated using key performance metrics, including overall accuracy (OA), precision (P), recall (R), and F-1 score.

2. Material and method

This study utilized the Google Earth Engine (GEE) Platform to perform the building extraction. GEE is a planetary-scale platform designed for Earth science and data analysis. The platform offers a web-based code editor equipped with a Javascript API for executing scripts. Consequently, it enables the use of cloud computing to quickly develop and perform complex geospatial workflows with ease.

The orthophoto of the study area had to be uploaded to the platform so as to perform the building extraction using the GeoBIA approach. The dataset was subsequently imported into the code editor, followed by the performance of the analysis. The workflow involved generating RGB VIs using the individual bands of the orthophoto, creating the datasets of each VI from the bands of the original orthophoto, adding each VI and the RGB image, performing an object-based segmentation on each dataset, selecting features for building and non-building classes, and training and testing a machine-learning classifier on the selected features.

The Randon Forest (RF) machine-learning classifier was adopted for this study, whereas the simple linear iterative clustering (SLIC) algorithm was exploited for the segmentation step. Figure 1 depicts the workflow adopted for this study.

2.1. Study area and dataset

The New Mankessim community is within the administrative jurisdiction of the Tarkwa Nsuaem Municipal Assembly, located approximately 19.30 kilometres southwest of the municipal capital, Tarkwa, in the Western Region of Ghana. From a geographical perspective, the community is positioned at latitude 5°5' 29.45" N and longitude 2°6' 4.70" W, nestled at an average altitude of 55 meters above mean sea level. In response to the evolving mining dynamics, a prominent mining company operating in the region initiated a resettlement program to relocate the community members from their previous dwellings to the current

location. Consequently, a well-planned layout is a notable feature of the New Mankessim community. In addition, a consistent architectural style marks this layout and reflects a cohesive and deliberate approach to urban development.

Aerial images of the area were acquired using a Phantom 4 UAV. These images were processed using Agisoft Metashape Pro photogrammetric software to obtain a georectified image of the community. Figure 2 depicts the georectified image of the study area.

2.2. Geographic object-based image analysis (GeoBIA)

GEOBIA is an image analysis approach that is commonly applied to VHR remote sensing data. It serves various purposes, such as land-cover mapping and identifying specific geographic objects like buildings, cars, and trees [35-36]. The workflow of the GeoBIA approach includes image segmentation, feature extraction, image classification, and accuracy assessment. These steps were all implemented using the Google Earth Engine platform.
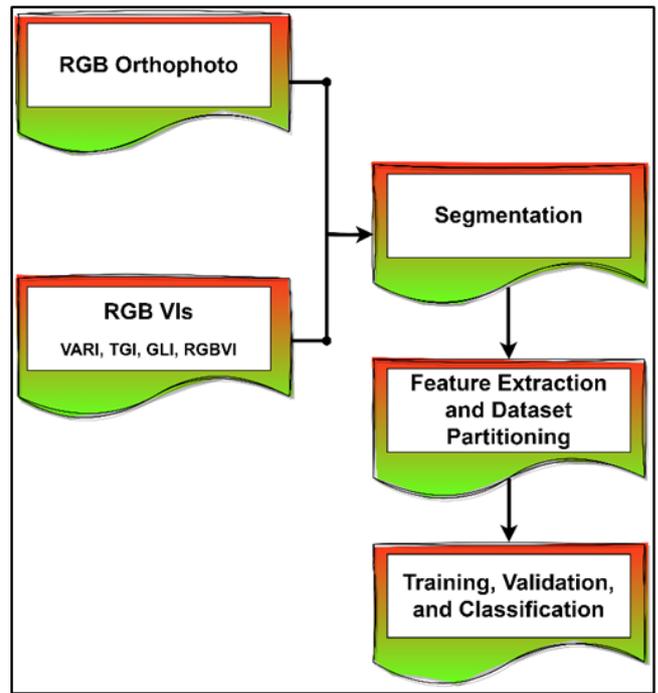

Figure 1. Methodological workflow.

Figure 2. UAV image of New Mankessim.

2.2.1 Image segmentation

This is the first step of geographic object-based image analysis, and it involves segmenting images into homogenous objects [23]. Image objects are groups of neighbouring pixels representing objects within the orthophoto with similar spectral and spatial attributes. There are various methods for performing image

segmentation; however, the simple linear iterative clustering (SLIC) algorithm proposed by Achanta et al. [37] was utilized. SLIC is a seed-based clustering technique that uses a modified k-means clustering strategy to create highly efficient superpixels. Unlike prior methodologies, SLIC excels in preserving boundaries while offering improved speed and memory efficiency.

It also enhances segmentation performance and can be extended for super voxel generation. This method carefully incorporates considerations of both color homogeneity and shape uniformity, achieving a well-balanced trade-off between these aspects [38]. To obtain optimum and homogenous image objects using SLIC, several parameters, such as compactness, seed size, and grid type, need to be defined.

However, due to the vast search space of the parameters, choosing the optimum combination of parameters is challenging. If the parameter combination is not carefully chosen, it can result in under-segmentation or over-segmentation. Thus, there is a high chance of low performance when the wrong parameter combination is used [25]. For this work, the parameters were determined using a trial-and-error approach, adopted according to the approach in [39, 40].

2.2.2 Feature extraction and selection

This step involves capturing substantial information from the image segments' objects to characterize features within an image. Commonly extracted features include spectral information, texture, shape, size, and contextual relationships. The most relevant features are subsequently selected to contribute to and optimize the computational efficiency of the subsequent classification process [3].

The spectral attributes, primarily the mean values of the red, green, and blue (RGB) bands, were extracted for this study. In addition, the mean of the spectral indices, that is, the green leaf index (GLI), red-green-blue vegetation index (RGBVI), visual atmospherically resistant index (VARI), and triangular greenness index (TGI), were extracted. These mean spectral values were chosen as the primary features for building extraction and further characterization because they capture colour information for distinguishing building objects from other urban features.

A total of 916 samples, 456 representing buildings and the remaining non-building objects, were selected to train the machine learning classifier. The samples were randomly divided into training (80%) and validation or test (20%) sets to facilitate model training and evaluation. This partitioning ensures that the machine learning classifier is robust and reliable, permitting effective learning. Moreover, it enables assessing the model's performance in distinguishing between building and non-building classes.

2.2.3 Classification

The step involves using a machine learning classifier to classify the segments into respective classes. For this research, the random forest (RF) classifier that Breiman [41] proposed was employed to classify the selected features as either buildings or non-buildings. RF is an ensemble machine learning algorithm that combines multiple decision trees to make predictions. Each tree in the forest is trained on a different subset of the data with bootstrapping and random feature selection. The final prediction is determined by a majority vote or averaging of individual tree predictions, making it robust, accurate, and less prone to overfitting, making it robust and effective in handling complex classification tasks [42, 43].

For this research, the RF classifier was trained using the 769 training samples, with the mean values of each dataset serving as input features. Like the SLIC algorithm, RF also has several parameters that must be fine-tuned for optimum classification. These include the number of trees, variablesPerSplit, minLeafPopulation, bagFraction, and maxNodes seed. Choosing a prejudiced parameter can result in overfitting or underfitting. Other than the number of trees set to fifty (50), default values were maintained for all the other parameters.

Subsequently, the performance of the trained RF model was assessed using the test data. This was to ascertain how the RF model fared on unseen datasets, prior to classifying the entire dataset.

The final step was to assess the binary raster produced by each dataset. As such, the classified datasets were exported, and an evaluation was performed using ArcMap. The create accuracy assessment points and update accuracy assessment points functions in the spatial analyst toolbox were adopted for this approach. Using the create accuracy assessment points function, the equalized stratified random sampling technique was used to generate 1500 data points each for the building and non-building classes. The ground truth mask, generated from the digitized building polygons, was used as the target field for this operation.

Subsequently, the update accuracy assessment points function was utilized to generate reference points for each classified dataset. These reference points were used to assess the validity of the classified images with regard to the ground truth mask.

2.3 Evaluation metrics

A comprehensive validation approach was adopted to assess the accuracy of the building classification. The trained RF classifier was applied to the validation data to classify buildings and non-buildings. The results were then compared with ground truth data to evaluate classification performance. This evaluation was based on four metrics: overall accuracy, precision, recall, and F1-score, all of which were computed using a confusion matrix.

Recall is a metric that quantifies the proportion of building image objects that were successfully classified as buildings. It measures how effectively the proposed methodology captures all existing buildings within the validation dataset. Conversely, precision offers insights into the correctness of the approach by elucidating the probability that a detected structure is indeed a building [33].

F1-score is a metric that considers precision and recall, thereby providing a balanced assessment of the

approach's performance. It is an essential metric when there is an imbalance between positive and negative instances in the dataset.

Overall accuracy is a metric that measures the ratio of correctly classified building instances to the total number of building instances in the dataset. Equations (1) to (4) give the mathematical formulations for the evaluation metrics.

$$\text{Recall} \qquad R = \frac{TP}{TP + FN} \qquad (1)$$

$$\text{Precision} \qquad P = \frac{TP}{TP + FP} \qquad (2)$$

$$\text{Overall Accuracy} \qquad OA = \frac{TP + TN}{TP + FP + TN + FN} \qquad (3)$$

$$\text{F}_1\text{-score} \qquad F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (4)$$

In Equations (1-4), TP represents correctly identified building segments, TN indicates correctly identified non-building segments, FP represents non-building segments erroneously classified as buildings, and FN denotes building segments incorrectly classified as non-building.

2.4 RGB-vegetative indices

Most consumer-grade UAVs are equipped with RGB cameras. RGB vegetation indices (VIs) are derived through mathematical equations applied to two or more spectral bands to emphasize specific aspects of vegetation greenness, thereby facilitating the identification of distinctive vegetation features within the imagery. This is possible as the digital number values of each band can be used to compute the pixel value in the RGB image. Several RGB VIs have since been created and developed [34, 44]. The RGB VIs utilized in this research are depicted in Figure 3, and their respective equations are in Table 1.
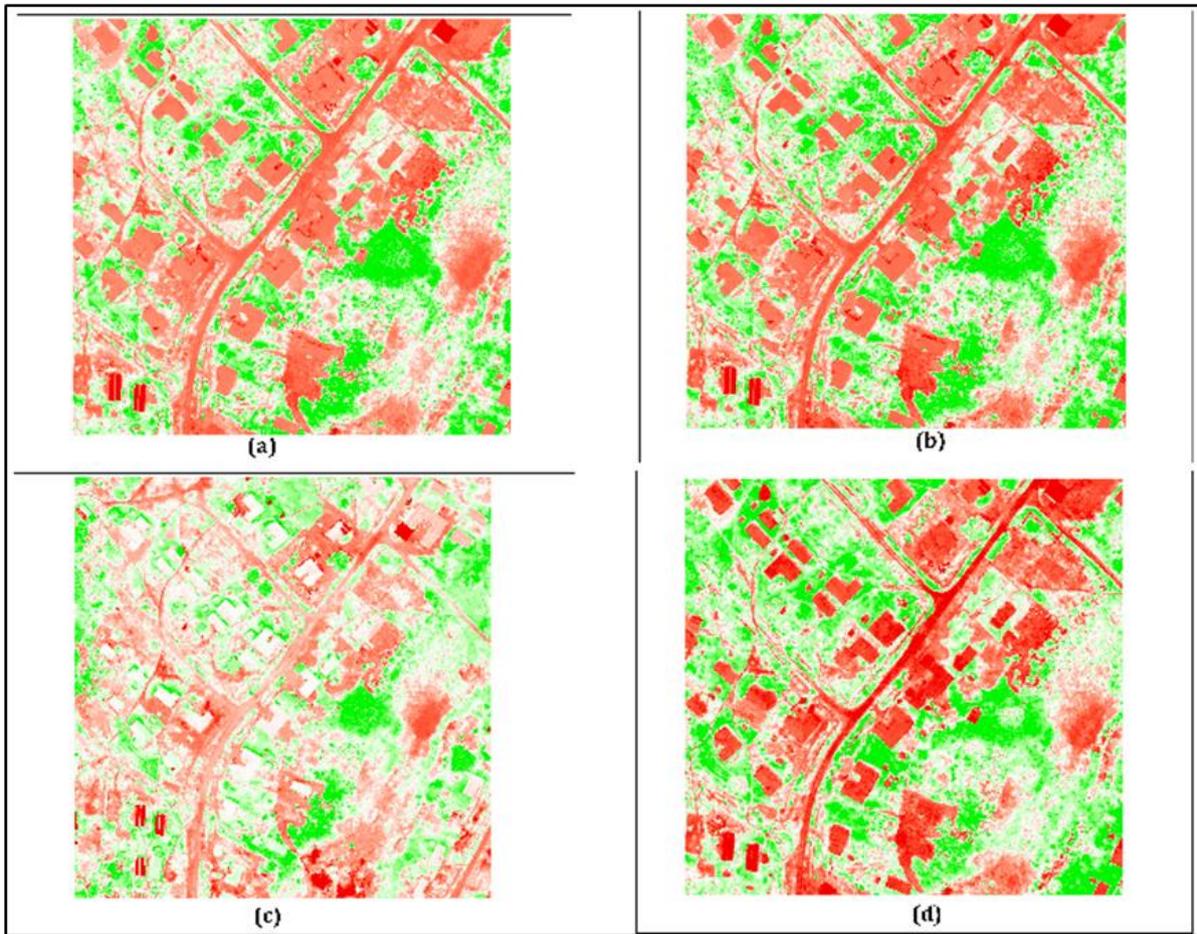

Figure 3. RGB VIs (a)GLI, (b)RGBVI, (c)VARI, and (d)TGI.

Table 1. RGB-VIs Utilized.

VI	Formula	Reference
Green Leaf Index	$GLI = \frac{(2 \times Green) - Red - Blue}{(2 \times Green) + Red + Blue}$	[45]
Red-Green-Blue Vegetation Index	$RGBVI = \frac{Green^2 - Blue \times Red}{Green^2 + Blue \times Red}$	[46]
Visual Atmospherically Resistant Index	$VARI = \frac{Green - Red - Blue}{Green + Red + Blue}$	[47]
Triangular Greenness Index	$TGI = Green - (0.39 \times Red) + (0.61 \times Blue)$	[48]

3. Results

This study generated five distinct datasets by combining the RGB VIs with the UAV-RGB image. These composite datasets included RGB with GLI, RGB with RGBVI, RGB with VARI, RGB with TGI, and RGB with all indices. Subsequently, each combination was used to train and validate the Random Forest (RF) classifier using selected spectral information.

The evaluation results obtained for each combination are detailed in Table 2. These results were derived from the 147 test datasets using the evaluation metrics, and these findings provide insight into how the

random forest model performed on the test datasets. A visual representation of Table 2 is presented in Figure 4. The confusion matrix from the validation process using the test data is illustrated in Figure 5.

The quantitative assessment derived from the classification result for each dataset is outlined in Table 3. These findings give insight into how each classified output corresponds with the ground truth. Figure 6 is a bar chart providing a visual representation of Table 3. The confusion matrix for the post-classification assessment is depicted in Figure 7, and Figure 8 presents the classification results obtained by each dataset.

Table 2. RF performance on test dataset for various combinations.

Dataset	Metric			
	OA	P	R	F1
UAV-RGB only	0.9565	0.9643	0.9529	0.9586
RGB + GLI	0.9632	0.9897	0.9411	0.9648
RGB +RGBVI	0.9660	0.9671	0.9671	0.9671
RGB + VARI	0.9799	0.9806	0.9806	0.9806
RGB + TGI	0.9714	0.9880	0.9535	0.9704
RGB + All Indices	0.9507	0.9570	0.9368	0.9468

Table 3. Classification performance of various datasets.

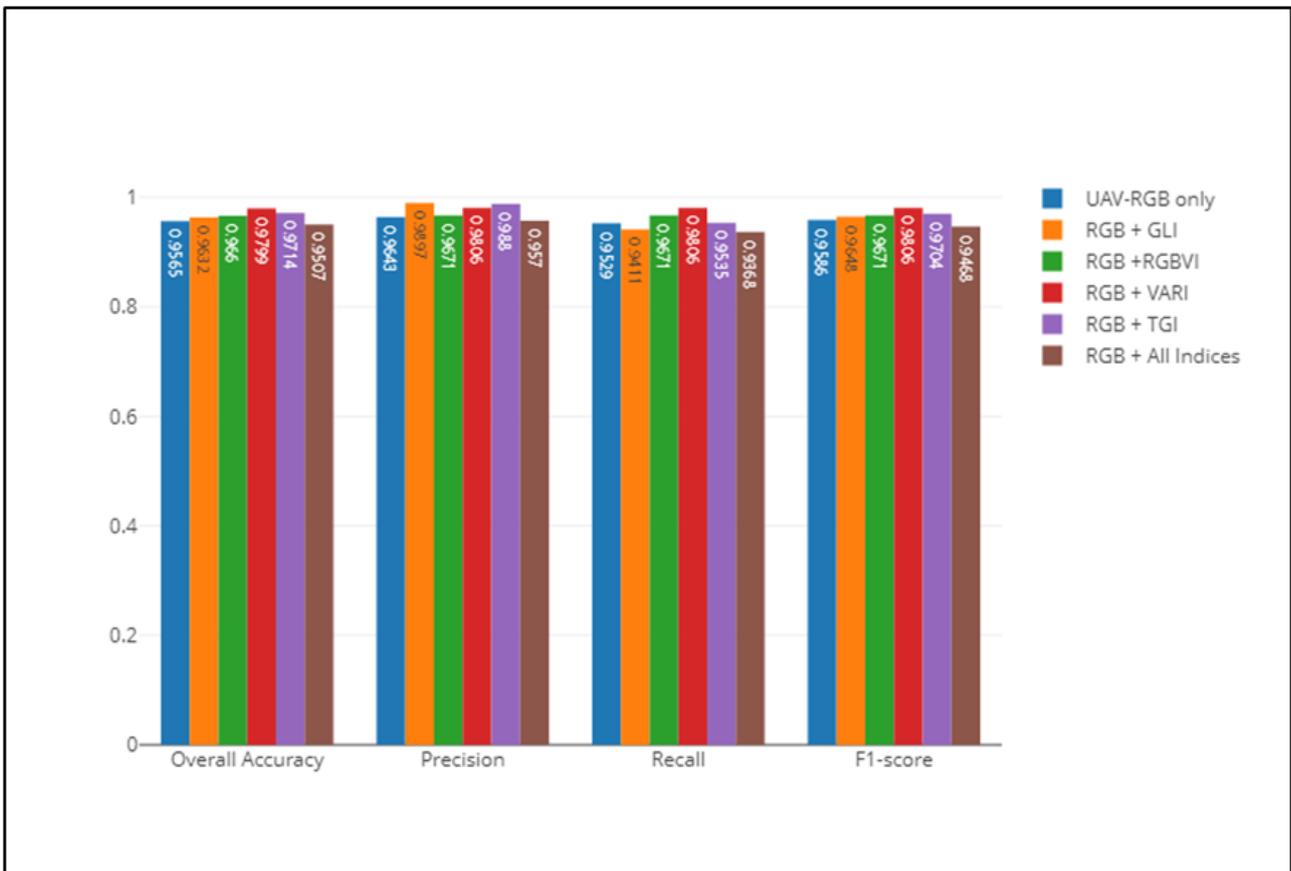| Dataset	Metric			
	OA	P	R	F1
UAV-RGB only	0.9553	0.9401	0.9727	0.9561
RGB + GLI	0.9673	0.9655	0.9693	0.9674
RGB +RGBVI	0.9647	0.9491	0.9820	0.9653
RGB + VARI	0.9603	0.9429	0.9800	0.9611
RGB + TGI	0.9657	0.9580	0.9740	0.9660
RGB + All Indices	0.9660	0.9575	0.9753	0.9663

Figure 4. Plot of RF validation result.

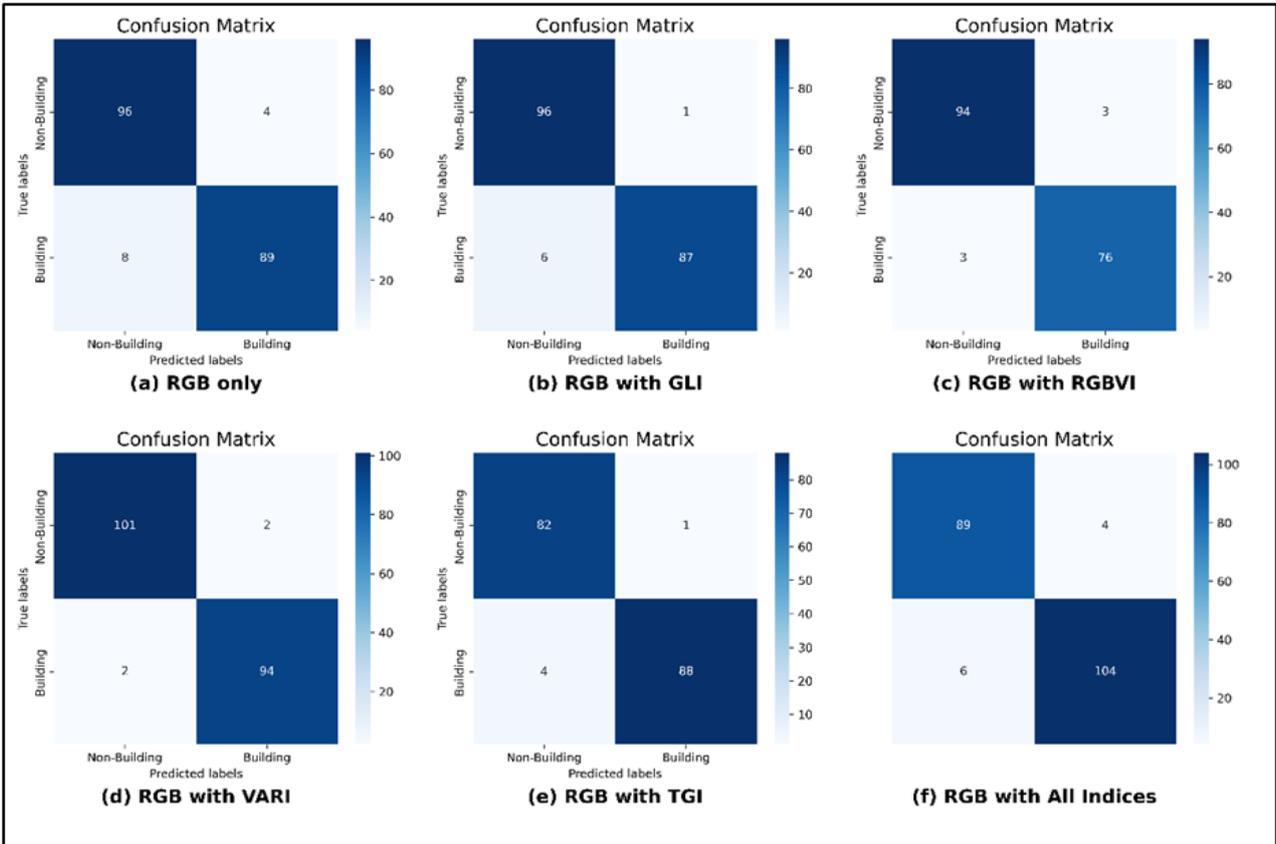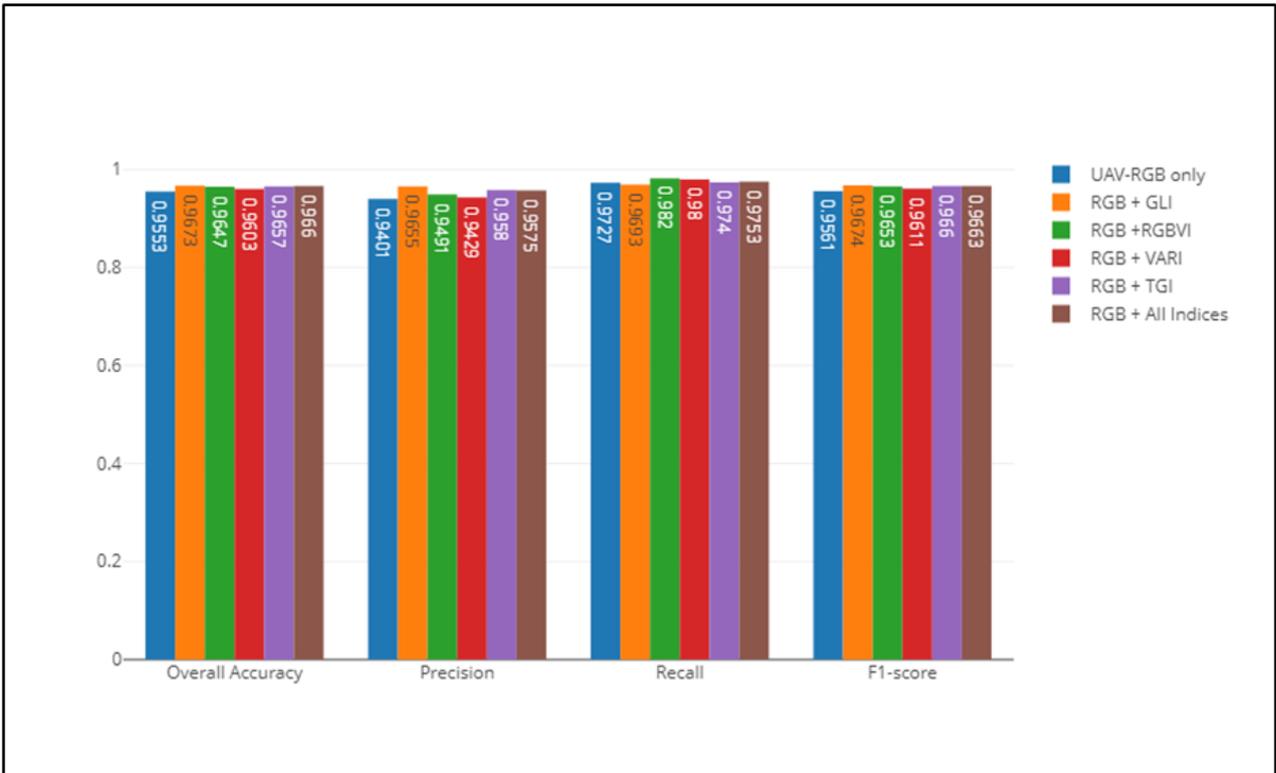**Figure 5.** Confusion matrix for test data validation.

Figure 6. Bar plot of classification validation.

4. Discussion

From Table 2, which is the validation result based on the test data, it is observed that the combination of RGB and VARI achieved the highest overall accuracy (0.9799), recall (0.9806), F1-score (0.9806), and third-best precision (0.9806). This dataset improved the performance of the approach in classifying building instances, attaining a mean improvement of 1.9975.

The RGB and TGI combination achieved the second-highest OA of 0.9714, the highest precision and F1-score of 0.9880 and 0.9704, respectively, and the third-best recall value of 0.9368, amounting to an average improvement of 0.6900.
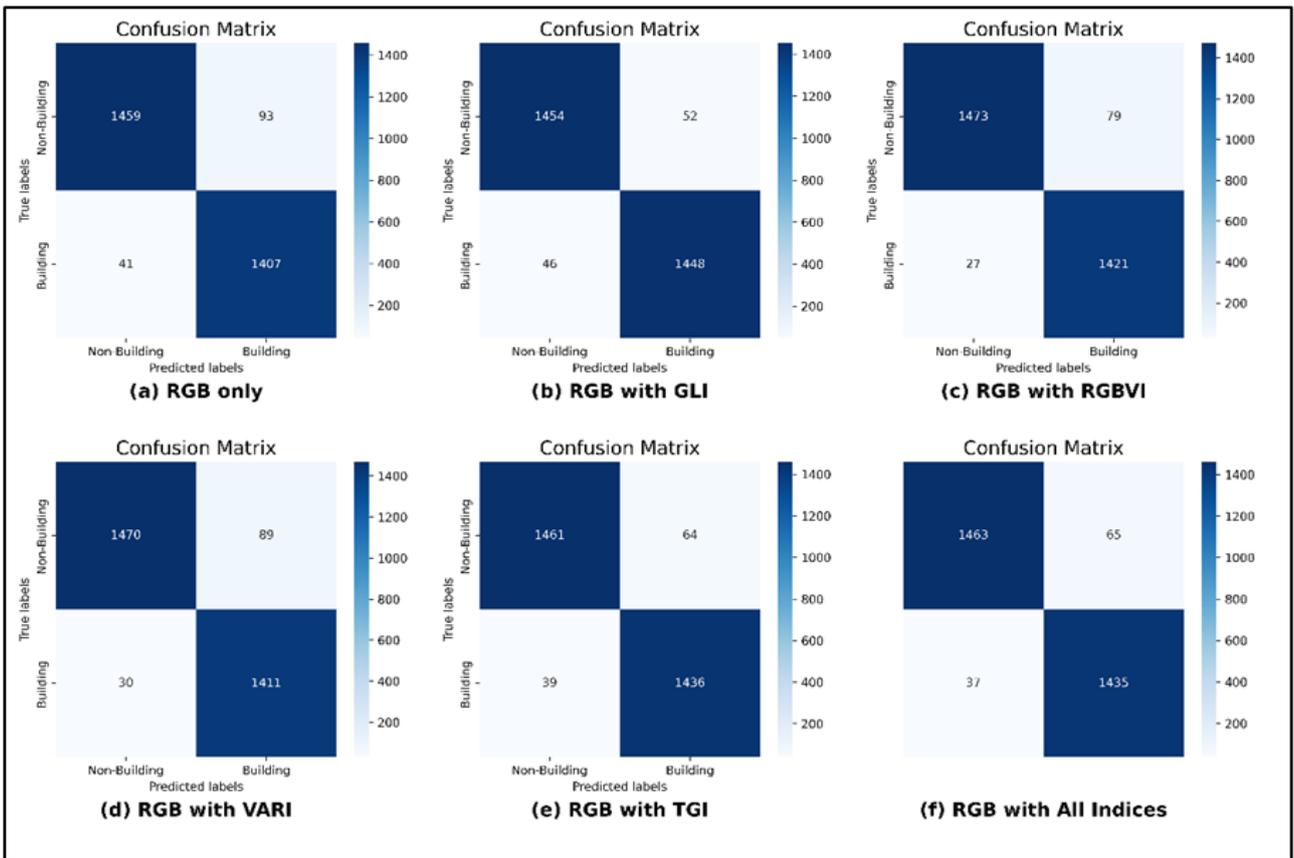
Figure 7. Confusion matrix for classification result validation.
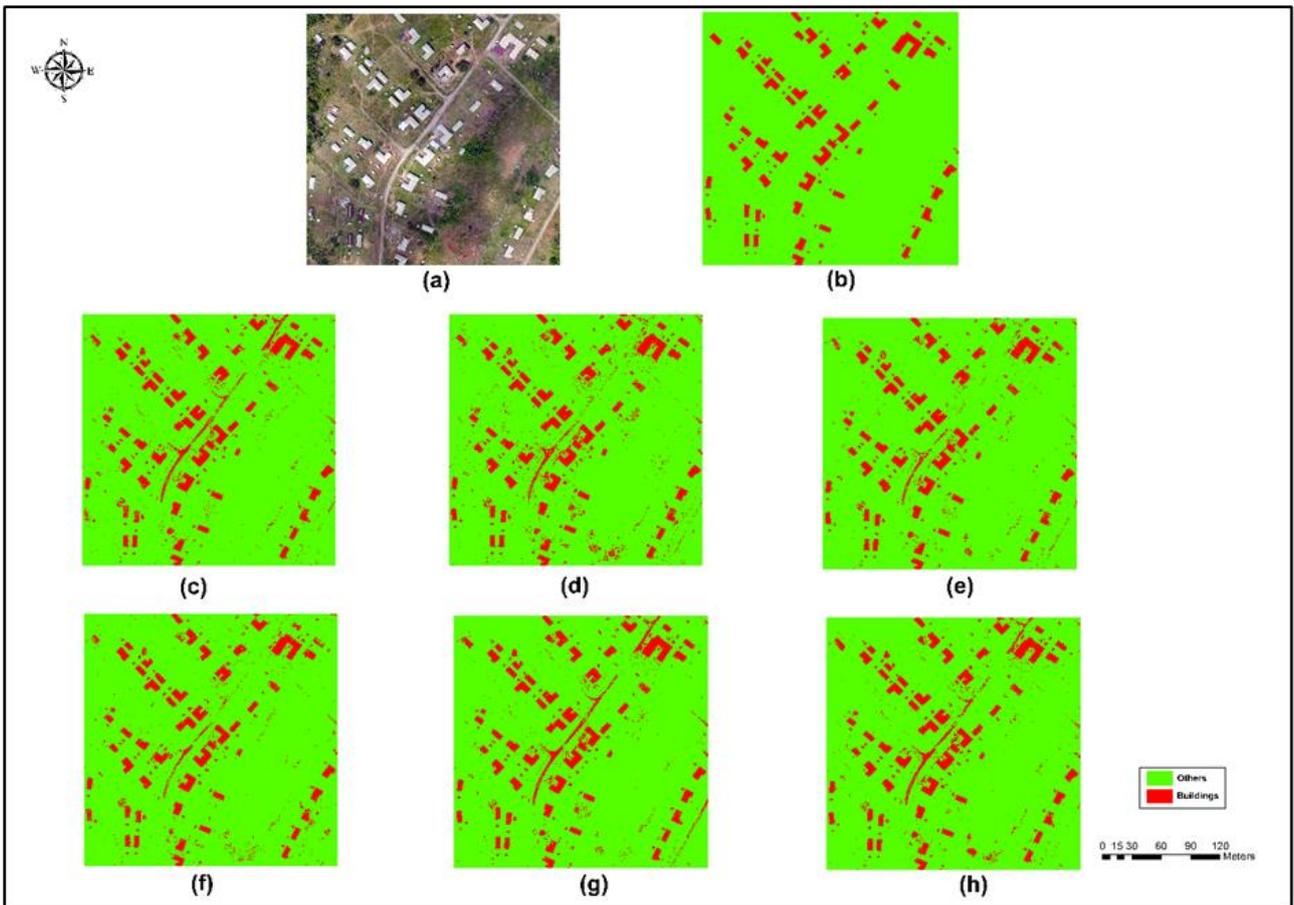

Figure 8. Building Extraction Results RF Classifier (a) UAV Image, (b) RGB + GLI, (c) RGB + RGBVI, (d) RGB + VARI, (e) RGB + TGI, and (f) RGB + All Indices.

The combination of RGB and RGBVI had the second-best recall (0.9671), third-best OA (0.9660), and F1-score (0.9671) values, and the fourth-best precision (0.9671). Thus, combining the RGBVI with the RGB orthophoto improved the classification approach by an average of 0.7075.

The RGB and GLI combination achieved the highest precision (0.9890). Notwithstanding, it ranked fourth in terms of overall accuracy (0.9632) and F1-score (0.9648) and fifth in recall (0.9411), which was lower than the standalone RGB combination. This combination attained a mean improvement of 0.6625.

The combination of all the RGB VIs and the RGB orthophoto had the least values for all the metrics, attaining an OA of 0.950, precision of 0.9570, recall of 0.9368, and F1-score of 0.9468. This result implies that combining all the VIs with RGB orthophoto decreased the performance of the building extraction approach by an average of -0.3975.

When considering only the RGB orthophoto, the overall accuracy, precision, recall, and F1-score values were 0.9565, 0.9643, 0.9529, and 0.9586, respectively, with the recall value ranking third, surpassing the RGB with GLI and RGB with all indices combinations.

The quantitative assessment of the classification result in Table 3 indicated that the RGB and GLI combination attained the best classification compared to the ground truth. The RGB and GLI attained the highest overall accuracy (0.9673), precision (0.9655) and F1-score (0.9674). Interestingly, this combination attained the lowest recall value of 0.9693.

Combining all the VIs with the RGB orthophoto achieved the second-best overall accuracy (0.9660) and F1-score (0.9663). This dataset combination also attained the third-best scores for precision (0.9575) and recall (0.9753). These scores contrast sharply to those attained for the test sample, where the dataset attained the lowest scores.

The third-best classification result was obtained by combining RGB and TGI. This dataset obtained the second-best score for precision (0.9580), the third-best score for overall accuracy (0.9657), and the F1-score (0.9660). For recall, this dataset obtained the fourth-best score of 0.9740.

From Table 3, the RGB and VARI dataset had the fourth-best classification output. This combination had the second-best recall value of 0.9800 and the fifth-best values for overall accuracy (0.9603) and precision (0.9429). Notwithstanding, the dataset attained the lowest value of 0.9611 for the F1-score.

The RGB and RGVI dataset achieved the fourth-best values for overall accuracy (0.9647) and precision (0.9491). This dataset also gained the fifth-best result (0.9653) for F1-score and the lowest value (0.9820) for recall.

Lastly, the RGB dataset had the lowest classification performance. This dataset had the lowest values of 0.9553 and 0.9401 for overall accuracy and precision, respectively. Nevertheless, the dataset attained the fourth-best value for F1-score (0.9561) and the fifth-best value for recall (0.9727).

These two results underscore the contribution of the VIs in the building extraction pipeline. Nonetheless, it is

evident that the VIs had varying performances. From the test dataset, it is evident that the RGB and VARI dataset outperforms the other datasets in terms of overall accuracy (OA), precision (P), recall (R), and F1-score. It was also interesting to note that the RGB and all Indices dataset performed slightly worse than the UAV-RGB only dataset. This dataset attaining the worst performance on the test data can be attributed to the curse of dimensionality, where redundant spectral information from the VIs introduced noise and multicollinearity. Thus, the effectiveness of a classification model is hindered, and the chance of overfitting the training dataset is increased [49]. The result from the classification showed a slight difference from the testing sample results. While the RGB and VARI dataset still performs well, the RGB and GLI and RGB and all Indices datasets demonstrated an improved performance.

The qualitative maps in Figure 8 show that the datasets produced similar visual outputs. All the datasets could identify and extract the outlines of the buildings of within the image. However, just like most attempts of automation in building extraction, there is the need for some manual editing and revision [50]. From Figure 6, it is noticeable that there was confusion among building rooftops and impervious surfaces such as roads and soil surfaces. This cataclysm resulted in commission errors, leading to an inconsistent classification. This can be attributed to the VIs' strength in differentiating between vegetation and non-vegetation rather than among urban features. Figure 6 shows that all the datasets falsely classified a section of the central road as a building.

5. Conclusion

This study aimed to assess the contribution of four RGB VIs, GLI, RGBVI, VARI, and TGI, in improving building classification tasks from UAV imagery. To that aim, four datasets containing a combination of these VIs and RGB-UAV were created, and a GeoBIA approach was adopted to classify building features from these datasets. In addition, a fifth dataset was created by combining all the RGB VIs and the UAV image.

The experimental results highlight the advantages of integrating vegetative indices into building extraction from UAV-RGB imagery. The RGB and VARI dataset emerged as the top-performing combination, achieving the highest overall accuracy, precision, recall, and F1-score on the test dataset. However, it is worth noting that the RGB and GLI dataset stood out for its exceptional precision. This result was useful during the classification step, where the dataset attained the result. Thus, it produced a classification result that was similar to the ground truth mask.

In conclusion, integrating the RGB-based VIs into the GeoBIA classification pipeline significantly improved the accuracy of building extraction from UAV-RGB imagery. Nevertheless, the choice of VIs and their combination plays a crucial role in the performance of the extraction approach. The combined use of all indices does not consistently outperform individual indices, emphasizing the significance of selecting relevant combinations. Future research could focus on choosing the best combination of VIs that further enhance the performance

of building extraction. Also, other machine-learning classifiers can be adopted, and their performance can be assessed during the classification step.

Author contributions

Richmond Akwasi Nsiah: Conceptualization, Methodology, Investigation, Software and Scripting, Data Curation, Validation, Writing—original draft preparation, Writing—review and editing, **Visualization Saviour Mantey:** Conceptualization, Methodology, Investigation, Validation, Writing—original draft preparation, Writing—review and editing, Visualization, Supervision **Yao Yevenyo Ziggah:** Conceptualization, Methodology, Investigation, Validation, Writing—original draft preparation, Writing—review and editing, Visualization, Supervision.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Schlosser, A. D., Szabó, G., Bertalan, L., Varga, Z., Enyedi, P., & Szabó, S. (2020). Building extraction using orthophotos and dense point cloud derived from visual band aerial imagery based on machine learning and segmentation. Remote Sensing, 12(15), 2397. https://doi.org/10.3390/rs12152397

2. Hu, Q., Zhen, L., Mao, Y., Zhou, X., & Zhou, G. (2021). Automated building extraction using satellite remote sensing imagery. Automation in Construction, 123, 103509.
   https://doi.org/10.1016/j.autcon.2020.103509

3. Li, J., Huang, X., Tu, L., Zhang, T., & Wang, L. (2022). A review of building detection from very high resolution optical remote sensing images. GIScience & Remote Sensing, 59(1), 1199-1225.
   https://doi.org/10.1080/15481603.2022.2101727

4. Dai, Y., Gong, J., Li, Y., & Feng, Q. (2017). Building segmentation and outline extraction from UAV image-derived point clouds by a line growing algorithm. International Journal of Digital Earth, 10(11), 1077-1097.
   https://doi.org/10.1080/17538947.2016.1269841

5. Temenos, A., Temenos, N., Doulamis, A., & Doulamis, N. (2022). On the exploration of automatic building extraction from RGB satellite images using deep learning architectures based on U-Net. Technologies, 10(1), 19.
   https://doi.org/10.3390/technologies10010019

6. Daranagama, S., & Witayangkurn, A. (2021). Automatic building detection with polygonizing and attribute extraction from high-resolution images. ISPRS International Journal of Geo-Information, 10(9), 606.
   https://doi.org/10.3390/ijgi10090606

7. Lin, Huertas, & Nevatia. (1994). Detection of buildings using perceptual grouping and shadows. In 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 62-69.
   https://doi.org/10.1109/CVPR.1994.323811

8. Jaynes, C. O., Stolle, F., & Collins, R. T. (1994, December). Task driven perceptual organization for extraction of rooftop polygons. In Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, 152-159.
   https://doi.org/10.1109/ACV.1994.341303

9. Chen, R., Li, X., & Li, J. (2018). Object-based features for house detection from RGB high-resolution images. Remote Sensing, 10(3), 451.
   https://doi.org/10.3390/rs10030451

10. Lu, T., Ming, D., Lin, X., Hong, Z., Bai, X., & Fang, J. (2018). Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network. Remote Sensing, 10(9), 1496.
    https://doi.org/10.3390/rs10091496

11. Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117, 11-28.
    https://doi.org/10.1016/j.isprsjprs.2016.03.014

12. Lefèvre, S., Weber, J., & Sheeren, D. (2007). Automatic building extraction in VHR images using advanced morphological operators. In 2007 Urban Remote Sensing Joint Event, 1-5.
    https://doi.org/10.1109/URS.2007.371825

13. Ahmadi, S., Zoej, M. V., Ebadi, H., Moghaddam, H. A., & Mohammadzadeh, A. (2010). Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. International Journal of Applied Earth Observation and Geoinformation, 12(3), 150-157.
    https://doi.org/10.1016/j.jag.2010.02.001

14. Yari, D., Mokhtarzade, M., Ebadi, H., & Ahmadi, S. (2014). Automatic reconstruction of regular buildings using a shape-based balloon snake model. The Photogrammetric Record, 29(146), 187-205.
    https://doi.org/10.1111/phor.12060

15. Huertas, A., & Nevatia, R. (1988). Detecting buildings in aerial images. Computer Vision, Graphics, and Image Processing, 41(2), 131-152.
    https://doi.org/10.1016/0734-189X(88)90016-3

16. Peng, J., & Liu, Y. C. (2005). Model and context-driven building extraction in dense urban aerial images. International Journal of Remote Sensing, 26(7), 1289-1307.
    https://doi.org/10.1080/01431160512331326675

17. Sirmacek, B., & Unsalan, C. (2008). Building detection from aerial images using invariant color features and shadow information. In 2008 23rd International Symposium on Computer and Information Sciences, 1-5.
    https://doi.org/10.1109/ISCIS.2008.4717854

18. Liow, Y. T., & Pavlidis, T. (1990). Use of shadows for extracting buildings in aerial images. Computer Vision, Graphics, and Image Processing, 49(2), 242-277.
    https://doi.org/10.1016/0734-189X(90)90139-M

19. Irvin, R. B., & McKeown, D. M. (1989). Methods for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Transactions

on Systems, Man, and Cybernetics, 19(6), 1564-1575. https://doi.org/10.1109/21.44071

20. Wu, G., Shao, X., Guo, Z., Chen, Q., Yuan, W., Shi, X., ... & Shibasaki, R. (2018). Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. Remote Sensing, 10(3), 407.
https://doi.org/10.3390/rs10030407

21. Kokeza, Z., Vujasinović, M., Govedarica, M., Milojević, B., & Jakovljević, G. (2020). Automatic building footprint extraction from UAV images using neural networks. Geodetski Vestnik, 64(04), 545-561.
https://doi.org/10.15292/geodetski-vestnik.2020.04.545-561

22. Norman, M., Shahar, H. M., Mohamad, Z., Rahim, A., Mohd, F. A., & Shafri, H. Z. M. (2021). Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. In IOP Conference Series: Earth and Environmental Science, 620(1), 012010. https://doi.org/10.1088/1755-1315/620/1/012010

23. Comert, R., & Kaplan, O. (2018). Object based building extraction and building period estimation from unmanned aerial vehicle data. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 4(3), 71-76.
https://doi.org/10.5194/isprs-annals-IV-3-71-2018

24. Aminipouri, M. (2009). Object-oriented analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar-Es-Salaam, Tanzania [Master's thesis, University of Twente].

25. Guo, Z., & Du, S. (2017). Mining parameter information for building extraction and change detection with very high-resolution imagery and GIS data. GIScience & Remote Sensing, 54(1), 38-63.
https://doi.org/10.1080/15481603.2016.1250328

26. 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.

27. Frishila, A. A., & Kamal, M. (2019). The effectiveness of spectral features for building extraction using geographic object-based image analysis (GEOBIA). The 40th Asian Conference on Remote Sensing (ACRS 2019), 1-10.

28. Hossain, M. D., & Chen, D. (2022). A hybrid image segmentation method for building extraction from high-resolution RGB images. ISPRS Journal of Photogrammetry and Remote Sensing, 192, 299-314.
https://doi.org/10.1016/j.isprsjprs.2022.08.024

29. Dornaika, F., Moujahid, A., El Merabet, Y., & Ruichek, Y. (2016). Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors. Expert Systems with Applications, 58, 130-142.
https://doi.org/10.1016/j.eswa.2016.03.024

30. Argyridis, A., & Argialas, D. P. (2016). Building change detection through multi-scale GEOBIA approach by integrating deep belief networks with fuzzy ontologies. International Journal of Image and Data Fusion, 7(2), 148-171.
https://doi.org/10.1080/19479832.2016.1158211

31. Davydova, K., Cui, S., & Reinartz, P. (2016). Building footprint extraction from digital surface models using neural networks. In Image and Signal Processing for Remote Sensing XXII, 10004, 187-196.
https://doi.org/10.1117/12.2240727

32. Li, Y., Zhu, L., Shimamura, H., & Tachibanab, K. (2010). An integrated system on large scale building extraction from DSM. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38, 35-39.

33. Singh, D., Maurya, R., Shukla, A. S., Sharma, M. K., & Gupta, P. R. (2012). Building extraction from very high resolution multispectral images using NDVI based segmentation and morphological operators. In 2012 Students Conference on Engineering and Systems, 1-5.
https://doi.org/10.1109/SCES.2012.6199034

34. Öztürk, M. Y., & Çölkesen, İ. (2021). The impacts of vegetation indices from UAV-based RGB imagery on land cover classification using ensemble learning. Mersin Photogrammetry Journal, 3(2), 41-47. https://doi.org/10.53093/mephoj.943347

35. Kucharczyk, M., Hay, G. J., Ghaffarian, S., & Hugenholtz, C. H. (2020). Geographic object-based image analysis: a primer and future directions. Remote Sensing, 12(12), 2012.
https://doi.org/10.3390/rs12122012

36. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274-2282.
https://doi.org/10.1109/TPAMI.2012.120

37. Liao, N., Liu, H., Li, C., Ren, X., & Guo, B. (2022). Simple linear ıterative clustering with efficiency. Advances in Intelligent Information Hiding and Multimedia Signal Processing, 1, 109-117.
https://doi.org/10.1007/978-981-19-1057-9_11

38. Zhang, H., & Zhu, Y. (2019). Kslic: K-mediods clustering based simple linear iterative clustering. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 519-529.
https://doi.org/10.1007/978-3-030-31723-2_44

39. Sibaruddin, H. I., Zulhaidi, H., Shafri, M., Pradhan, B., & Haron, N. A. (2018). UAV-based approach to extract topographic and as-built information by utilising the OBIA technique. Journal of Geosciences and Geomatics, 6(3), 103-123.
https://doi.org/10.12691/jgg-6-3-2

40. Norman, M., Shahar, H. M., Mohamad, Z., Rahim, A., Mohd, F. A., & Shafri, H. Z. M. (2021). Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. IOP Conference Series: Earth and Environmental Science, 620(1), 1-11. https://doi.org/10.1088/1755-1315/620/1/012010

41. Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.

42. Kumar, A., & Sinha, N. (2020). Classification of forest cover type using random forests algorithm.

In Advances in Data and Information Sciences: Proceedings of ICDIS 2019, 94, 395-402. https://doi.org/10.1007/978-981-15-0694-9_37

43. Xiao, Y., Huang, W., & Wang, J. (2020). A random forest classification algorithm based on dichotomy rule fusion. In 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), 182-185. https://doi.org/10.1109/ICEIEC49280.2020.9152236

44. Jiang, J., Cai, W., Zheng, H., Cheng, T., Tian, Y., Zhu, Y., ... & Yao, X. (2019). Using digital cameras on an unmanned aerial vehicle to derive optimum color vegetation indices for leaf nitrogen concentration monitoring in winter wheat. Remote Sensing, 11(22), 2667. https://doi.org/10.3390/rs11222667

45. Hunt Jr, E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. International Journal of Applied Earth Observation and Geoinformation, 21, 103-112. https://doi.org/10.1016/j.jag.2012.07.020

46. Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., ... & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. International Journal of Applied Earth Observation and Geoinformation, 39, 79-87. https://doi.org/10.1016/j.jag.2015.02.012

47. Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. Remote Sensing of Environment, 80(1), 76-87. https://doi.org/10.1016/S0034-4257(01)00289-9

48. Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto International, 16(1), 65-70. https://doi.org/10.1080/10106040108542184

49. Kaur, R., & Pandey, P. (2022). A review on spectral indices for built-up area extraction using remote sensing technology. Arabian Journal of Geosciences, 15(5), 391. https://doi.org/10.1007/s12517-022-09688-x

50. Tsai, Y. H., Stow, D., & Weeks, J. (2011). Comparison of object-based image analysis approaches to mapping new buildings in Accra, Ghana using multi-temporal QuickBird satellite imagery. Remote Sensing, 3(12), 2707-2726. https://doi.org/10.3390/rs3122707