RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2624

Weighted Hermitian Wavelet Multilayer Extreme Learning Machine for Building Detection Using Unmanned Aerial Vehicle Images

A. Franklin Alex Joseph [a], A. Manikandan [b], Umi Salma Basha [c], Shanmugapriya Velmurugan [d,*], Rajesh Natarajan [e], S. Subash Chandra Bose [f]

[a] Department of Electronics and Communication Engineering, Periyar Maniammai Institute of Science and Technology, Thanjavur, Tamil Nadu, India.
[b] Department of Computer Science and Engineering, Paavai Engineering College, Namakkal (Dt), Tamil Nadu, India.
[c] Department of Computer Science and Engineering, Jazan University, Gizan, Saudi Arabia
[d] Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.
[e] Department of Computing and Information Sciences, College of Computing and Information Sciences, University of Technology and Applied Sciences, A1-Aqr Shinas, 324, Oman
[f] Department of Computer Applications, East Point College of Higher Education, Bengaluru, Karnataka, India.
* Corresponding Author Email: vspriyaphd@gmail.com

DOI: https://doi.org/10.54392/irjmt2624
Received: 02-04-2025; Revised: 15-02-2026; Accepted: 27-02-2026; Published: 10-03-2026

Abstract: Robust building extraction is crucial for 3D building models in Unmanned Aerial Vehicle (UAV) images. Automatic building extraction is in extremely high demand due to the productivity gain. Conventional building detection methods were time-consuming, costly, and highly complex for human experts. Several DL techniques were developed for building detection. But the accuracy was not enhanced. To solve this issue, the Weighted Hermitian Wavelet Multilayer Extreme Learning Machine (WHWMELM) technique is proposed with multiple layers for building detection in UAV images. The aim of the proposed method is to accurately detect building objects with maximum accuracy and minimal time. First, the input layer obtains UAV aerial input images, and the Weighted Myriad Filtering model is utilized for eliminating noise in the first hidden layer. The Hermitian multi-wavelet transform uses the next hidden layer to extract color, texture, and shape features. Later, the Schutz Feature Matching Coefficient classifies UAV images into buildings and non-buildings with higher accuracy and less time. The methodology has demonstrated promising results on a demanding benchmark dataset, significantly reducing time.

Keywords: Unmanned Aerial Vehicle, UAV Aerial Images, Building Detection, Multilayer Extreme Learning Machine.

1. Introduction

The automatic detection of buildings has always been a hot topic of study in the field of aerial image processing. In modern years, UAVs have developed rapidly and provided high-resolution aerial images to detect and recognize surface objects. DL has attained immense attention in areas of speech recognition, image recognition, information retrieval, etc. CNN Branch Reinforce Transformer, termed BranTNet, was developed in [1] for finding the straw fires in UAV Aerial Images. With this, environmental protection and health preservation were ensured. However, the detection accuracy was lower. A rural-created building in UAV was discovered in [2] by the Multiscale Fusion Network. Developed network addressed the problems related to the diversity of UAV images. However, time consumption was not reduced.

An Improved Lightweight Network was introduced in [3] to extract building footprints from UAV images. Fuse feature maps were used to get building. However, the extraction rate was not improved. A deep learning model was described in [4] to robotically find sugarcane crop lines using UAV images. CNN was employed to partition the images. Identified lines were enhanced with a refinement procedure. But, the noisy pixel was not eradicated to enhance the image quality.

Deep learning was employed in [5] with UAVs. Nevertheless, the feature extraction was not performed. An approach for detecting weeds in Chinese cabbage depended on the deep learning model described in [6]. It used UAV images as sample images. However, it failed to perform preprocessing.

A CNN-based model for identifying the individual tree was designed in [7] by using visible-thermal UAV images. Trees were discovered at thermal image based on brightness temperature variation. But, the precision and recall were not concentrated. UAV TIR object detection approach was described in [8] for UAV.

Features extracted with CNN via YOLO. Ground-based TIR images and videos captured with FLIR cameras. However, object detection time was not reduced.

A novel scheme to eliminate soil contribution from images that enhances the accuracy of maize at all development epochs was explained in [9].The ensemble learning model showed good strength in computing maize LAI at diverse growth stages. But, the time was higher. Cotton plants were discovered by DL through RGB images acquired by a UAV designed in [10]. The method was designed based on the YOLOv3 for the detection of VC plants. However, the accuracy was not enhanced.

Traditional building detection approaches were designed for classifying objects. The above existing methods some limitations are identified such as the lesser accuracy, higher time consumption, failed to consider error rate, and poor image quality. To overcome the issue, proposed WHWMELM is introduced for building detection.

The major contributions of the WHWMELM technique are provided as follows.

- The proposed WHWMELM technique is developed to accurately detect building images with preprocessing, feature extraction, and classification.
- The Weighted Myriad Filtering model is applied to perform preprocessing in the WHWMELM technique. It determines and eliminates the noisy pixels to get better-quality images. In this way, the PSNR is improved and the MSE is reduced.
- Hermitian multi-wavelet transform is utilized in the WHWMELM technique to decompose images into diverse frequency bands for extracting the features such as texture, color, and shape. With this, the classification time is minimized.
- The Schutz Feature Matching Coefficient approach is employed in the WHWMELM technique for classification. The similarity between two features is measured. An image into building objects or non-building objects is categorized. As a results, classification accuracy is enhanced.

Manuscript organized as given by. Related works are explained in Section 2. The UAV system presented in Sections 3 and 4 explains the research procedure with the following architecture diagram. Section 5 describes results and discussion. Section 6 concludes conclusion.

2. Literature Survey

To categorize plant communities of karst wetland, UAV images by spaceborne Jilin-1 multispectral images were used in [11]. Combined visible/multi-spectral UAV images using deep learning models were developed in [12] for object identification to broccoli development monitoring to obtain a visualized map.

In [13], evaluation amid Pixel-based of DL as well as Object-based of image examination was carried out depending on UAV images. For computing resources of private native forests, drone-based monitoring and image processing schemes were developed in [14]. The major purpose was to make competence for performing forest list remote with no relying on laborious meadow evaluations.

During preprocessing, feature extraction, and classification, Satin Bowerbird Optimization through DL for Food Crop Classification was developed in [15] using UAV images.

It used a bilateral filtering model to get preprocessed images. MobileNetv2 feature extractor was employed to extract the features. Also, classification using a Convolutional LSTM (ConvLSTM) model was applied to carry out crop categorization. But, the accuracy was not improved. A lightweight CNN architecture named EmergencyNet was developed in [16] to significantly classify the UAV images for emergency response/monitoring scenarios.

In [17], lightweight DL networks were developed to determine Derris trifoliate by little data. White grapes and berries from a UAV are discovered in [18] by Multi-object tracking and segmentation. PointTrack algorithm was employed to find the grape bunches. Spatial Embeddings were considered to detect berries. A high-resolution hybrid network was designed in [19] for object detection in UAV images. Detection of maize tassels was carried out in [20] finds maize tassels.

A weighted ensemble transferred UNet-based model (WETUM) was developed in [21] to generate a building damage map (BDM). Mask Region-Convolutional Neural Network (Mask R-CNN) was introduced in [22] for obtaining automatic building detection. But, the classification was not performed. LUD-You Only Look Once (YOLO) was investigated in [23] for handling the issue of image feature degradation. Computational cost and memory were reduced by using C2f-BiLevel Routing Attention (C2f-BRA) feature extraction. Pruning operations were carried out to decrease the complexity of the network. However, the accuracy was not enhanced. Buildings were extracted in [24] by employing Res-Unet. Nevertheless, the memory was higher. U-Net was discussed in [25] for enhancing image quality with semantic segmentation. However, the huge dataset was not concentrated.

3. UAV scheme

3.1 Why Develop a UAV Platform for Building Detection?

Identification of objects (i.e., buildings) as recorded on imagery obtained has a general process inside digital photogrammetry. Nevertheless, meeting timely as well as trouble-free operation constraint to remote sensing stage is in demand. Fixed-wing aircraft were employed as well as they needed a professional crew of pilots, navigators, as well as could operate in fine weather. Two demands determinefor timely information. 1) Presented another airborne platform2) camera system, and enhanced data processing with growth of dedicated software packet as well as workflow.

For capturing images of small areas, UAVs prepared by image sensors were suited. UAVs do not necessitate well-skilled operation and this leads to easier staff employment, lower costs, as well as work disposition. UAVs do not have similar strict operational control with external authorities, comprises Peking University as well as the Chinese Academy of Survey and Mapping. The last analysis resulted in a palette of products during recently developed components that numerous UAV platforms and cameras are integrated for the image acquisition system.

4. Building Detection by WHWMELM Technique

Buildings have presented as a basic part to forming a town.

Hence, modern building maps are imperative with different applications, including urban mapping as well as expansion investigation. Detection of building footprints is subject of intense study, by design of DL. Therefore, building detection using the WHWMELM technique from UAV images is introduced in this research. This technique takes advantage of simple, elastic, and low-cost UAV aerial imagery. The workflow of the proposed WHWMELM technique to fulfill the objectives of this research is illustrated in Figure 1. This technique mainly consists of three stages which are introduced as follows in detail.

Figure 1 shows the process proposed WHWMELM technique which consists of three phases such as preprocessing, feature extraction, and classification. The UAV images '$I_1, I_2, I_3, \ldots . I_n$' are collected from the dataset at different times for building detection. With the input images, the research procedure is explained as given below.

4.1 Preprocessing

To begin with, the proposed WHWMELM technique performs image preprocessing. The preprocessing step involves image resizing and noise removal while preserving the edges and enhancing the quality, resulting in improved detection accuracy. In the proposed research, the preprocessing step is implemented to achieve better-quality images for classification.

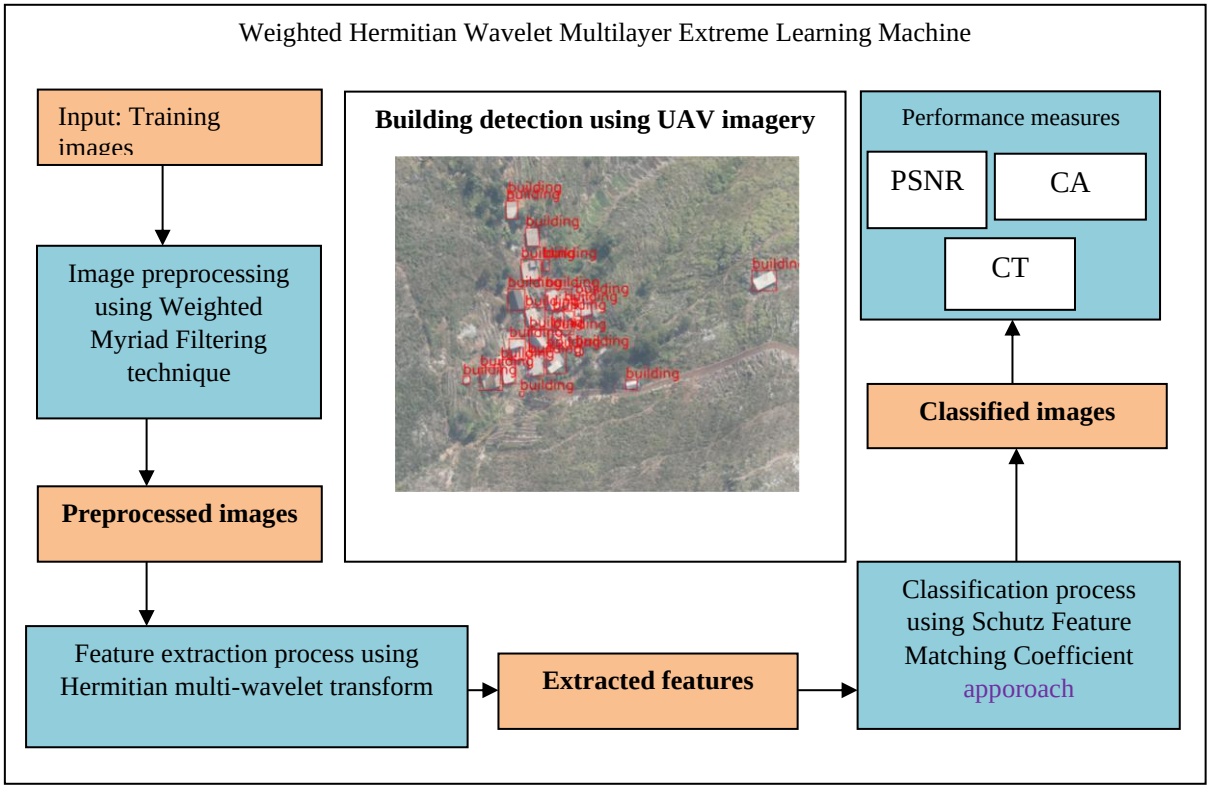

Figure 1. Flow process of Methodology for Building Detection

In the WHWMELM technique, preprocessing is performed using Weighted Myriad Filtering in the first hidden layer to take away noise from the image. To determine the noisy pixels in the given image, the filtering technique considers the maximum likelihood value. The UAV images taken from the input dataset are described as '$I_1, I_2, I_3, \ldots . I_n$'. The pixels in the input image are referred to as '$\sigma_1, \sigma_2, \sigma_3, \ldots . \sigma_m$'. Here, the pixels in the input images are related in the filtering kernel window through the representation of rows (i) and columns (j).

The center pixel is considered in the proposed weighted myriad filtering. It finds the maximum likelihood position evaluation among the center and neighboring pixels via a given mathematical formula.

$$O_g = arg\,max\left(w * \left|\sigma_{ij} - \sigma_k\right|\right) \quad (1)$$

In the above equation (4.1), O_g is a filtered output, $\arg max$ indicates an argument of the maximum function σ_{ij} denotes a center pixel and σ_k denotes a neighboring pixel, $w$and refers to a weight. The weight function in the filtering model helps to handle the blurring in input images. From the above equation, pixels that vary from the maximum likelihood of nearby pixels are identified. It's eradicated to get better-quality input images.

4.2 Data augmentation

Data augmentation is a vital task in building detection, land cover classification, and change detection for small datasets. Data augmentation is the process of creating new data (images) to produce additional information from limited data. The main goal of data augmentation is to enhance the volume, quality, and variety of training images. In our work, data augmentation is employed to generate a large number of UAV images. Also, the generalization and deep learning performance are enhanced.

4.3 Feature Extraction

In the second hidden unit, WHWMELM uses feature extraction with less time for building detection by using the Hermitian multi-wavelet transform. Feature extraction involves a critical algorithm for denoting the visual content of an input image. It denotes the original image in terms of a minimized form. Morlet wavelet transform and the Hermitian multi-wavelet transform are employed to decompose images into different frequency sub-bands. This decomposition is applied in different image processing tasks such as image compression, denoising, and feature extraction. However, the accurate time and frequency localization is a challenging issue. Hence, the Hermitian multi-wavelet transform is selected to perform feature extraction for separating the images into four sub-bands such as Low-Loe (S_{ll}), Low-High (S_{lh}), High-Low (S_{hl}),, and High-High (S_{hh}).

After denoising, the high-frequency images are targeted and extracted. Less high-frequency images are discarded. In this way, the time is reduced. Multiple scaling and wavelet functions are employed in the Hermitian multi-wavelet transform for examining preprocessed images with higher feature extraction performance. The process of feature extraction is shown below Figure 2.

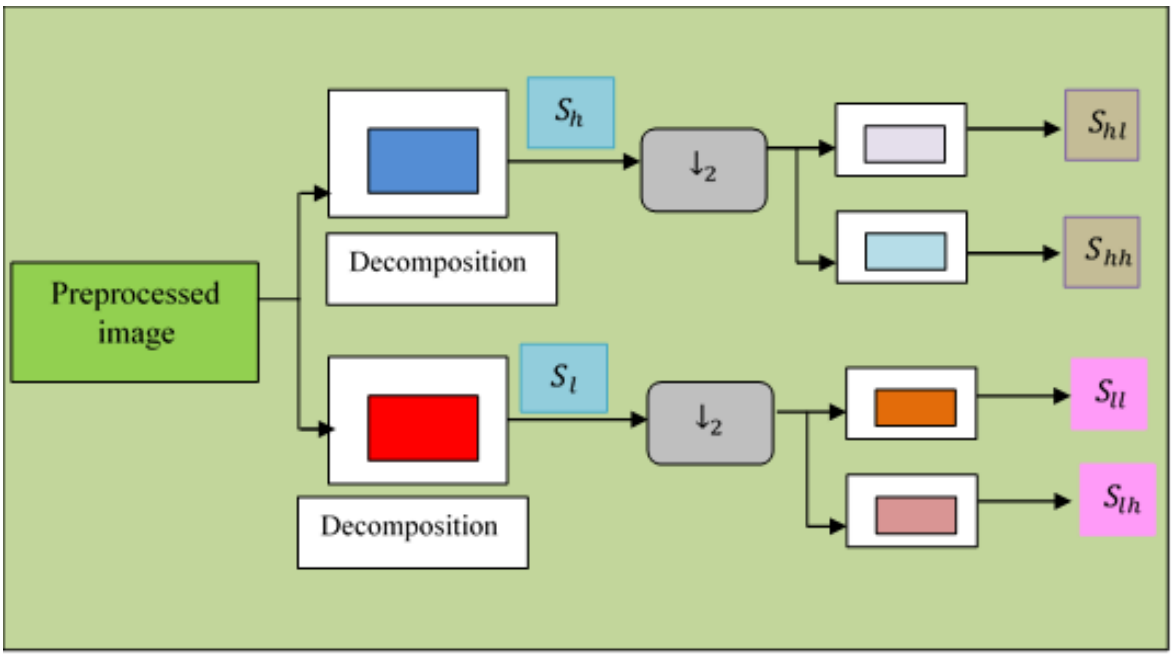

Figure 2. Hermitian multi-wavelet transform-based image decomposition.

As shown in the figure, preprocessed images are given as input to the transformation. Then the Hermitian multi-wavelet transform is applied to decompose the image into different sub-bands with low and high frequencies. Upon feature decomposition, features are obtained. In the primary stage, input images are partitioned into low (S_l) as well as high (S_h). Horizontal as well as perpendicular paths of the image are divided by the Hermitian multi-wavelet transformation. Then, next successive steps, down sampling carried out as well as output of every stage, create four sub-bands $S_{hl}, S_{hh}, S_{ll}, S_{lh}$. Once again the same process is executed on the sub-band to make the next decomposition level. The transformation is given by,

$$\varphi(t) = (2k)^{-0.5\,k} \alpha \delta_n(t)\,(e^{-\frac{1}{2k}t^2}) \quad (2)$$

$$\tau(t) = (2k)^{-0.5\,k} \alpha W_n(t)\,(e^{-\frac{1}{2k}t^2}) \quad (3)$$

Where, multi-scale function in time 't' indicates $\varphi(t)$, 'k' as well as 't' are integers, '$\tau(t)$' denotes a multi-wavelet function, '$\delta_n(t)$' refers to a multi-scale filter coefficient, refers multi-wavelet filter coefficient is '$W_n(t)$', The normalization coefficient is 'α'. From this, 4 sub-bands are acquired from preprocessed images. Upon image decomposition, for extracting feature from the imagefeature extraction is employed.

4.4 Schutz Feature Matching Coefficient-based Building Detection

Lastly, the proposed WHWMELM carries out the classification to determine the building object in the input images with the help of the Schutz Feature Matching Coefficient. The third hidden layer carries out classification. The layer receives extracted features from the input images as input. The feature matching coefficient approach is employed in computer vision to discover features among two images. It involves identifying distinctive features in each image, such as corners, edges, or other significant structures, based on their similarity. This process is essential for different tasks such as building detection, object recognition, image alignment, and 3D reconstruction. The aim of the feature matching is to recognize and compare corresponding features (like edges, corners, or textures) in two or more images. The Jaccard index and Schutz index are employed to compute the similarity between two features. But the Jaccard index only considers shared attributes (intersection or present). Contrary to the existing Jaccard index, Schutz index considers both shared and unshared features (both intersection or present and differences or absent). In our work, the Schutz Feature Matching Coefficient approach is chosen in the WHWMELM technique to find the similarity between features or the matching between features detected in different images. The similarity coefficient value is from 0 to 1. The value '1' indicates features perfect match; the building is detected. The value '0' indicates features do not match, the non-building is detected. Based on the similarity score, the accuracy of finding a building or non- non-building is achieved. Schutz Feature Matching Coefficient approach is formulated as given below.

$$\beta = \frac{E_F \cap T_F}{\sum E_F + \sum T_F - E_F \cap T_F} \quad (4)$$

Where 'β' denotes a Schutz Feature Matching Coefficient, E_F points outs extracted features from the input image (i.e., training image), and T_F point outs features from the testing image, $E_F \cap T_F$ denotes a mutual dependence among the features. The results of the Schutz Feature Matching Coefficient approach are varied between 0 to 1. Lastly, the outputs are obtained at the output layer for building detection. The classification results of the output layer are provided by,

$$\beta = \begin{cases} 1\,;\, building\ objec \\ 0\,;\, non-\ building\ object \end{cases} \quad (5)$$

The result of '$\beta = 1$' denotes that the training and testing features are matched and the building is detected. On the other hand, '$\beta = 0$ denotes features that are not matched where the building is not detected. From that, the building objects are accurately determined depending on the feature-matching process. Later, the error 'ε' is computed at the output layer according to the variation between predicted and actual results.

$$\varepsilon = \left|\varepsilon_x - \varepsilon_y\right|^2 \quad (6)$$

From (6), 'ε_x'denoted as squared dissimilarity among actual error, predicted error is 'ε_y'. If the proposed technique attains smaller errors, the process is stopped, and the buildings are correctly detected. The algorithmic process of building detection is described as follows.

Input: Dataset, UAV images $I_1, I_2, I_3, \ldots . I_n$

Output: Accurate building detection

Begin

1. Get number of UAV images $I_1, I_2, I_3, \ldots . I_n$ //input layer
2. **For** each input UAV image //hidden layer 1
3. Perform image preprocessing
4. Eliminate noise from images
5. **For** each preprocessed image //hidden layer 2
6. Apply Hermitian multi-wavelet transform
7. Carry out first level transformation obtain S_l and S_h
8. Execute second level transformation obtain $S_{hl}, S_{hh}, S_{ll}, S_{lh}$

9. **End for**
10. **For**every sub-band
11. Extract texture, color, shape features
12. **End for**
13. Forward extracted features to an hidden layer 3
14. **For** extracted features E_F
15. Measure Schutz Feature Matching Coefficient approach
16. **If** $(\beta = 1)$**then**
17. Building object is detected
18. **Else**
19. Non-building object is detected
20. **End if**
21. **End for**
22. **End for**
23. Measure error ε
24. Repeat the process until minimum error is attained

End

As explained in the above algorithm, building detection from UAV images is achieved using WHWMELM with higher accuracy and less time. This can be achieved by initially preprocessing the input images by eliminating the noisy pixels. Then, the more informative features are extracted to reduce the time for building object detection. Later, resultant features are classified via the Schutz feature matching coefficient approach to determine whether the object is a building or non-building.

5. Results and Discussion

The experimental evaluation of the proposed WHWMELM technique and BranTNet [1], Multiscale Fusion Network [2], and WETUM [21] are implemented in Python using UAV Small Object Detection Dataset taken from https://www.kaggle.com/datasets/sovitrath/uav-small-object-detection-dataset. This dataset comprises aerial imagery from UAVs for little object detection employing deep learning. The dataset includes 844 images from UAVs with 10 different classes. The total number of instances is 18,234. Each image is in high resolution. Starting from 1600 pixels in width, some even go up to 2700 pixels in width. The dataset is divided into training, validation, and test. Training (717 samples), Validation (84 samples), and Test (43 samples) utilized. 10 different classes are utilized, such as Building, ship, vehicle, prefabricated-house, well, cable-tower, pool, landslide, cultivation-mesh-cage, and quarry. The aim of the dataset is to precisely classify the object (building or non-building) in UAV images. For conducting the experiments, the number of UAV images in the range from 100-1000 was employed. The performance of the WHWMELM technique is evaluated in various metrics. Hardware and Software configurations and hyperparameter details are listed in Tables 1 and 2.

Table 1. Hardware and Software Configurations

Software requirements	Hardware requirements
O/S: Windows 10 and above	CPU Processor: Intel® Core™ i3 Processor
Language: Python	Hard drive capacity: 1TB
	Mouse: Logitech
	Keyboard: 110 keys enhanced
	RAM (Memory capacity): 8GB
	System type: 32-bit Operating system

Table 2. Hyperparameter Details of Proposed WHWMELM technique

S. No	Hyperparameters	Description
1	Number of layers used in Multilayer Extreme Learning Machine	One input layer, three hidden layers, and one output layer
2	Activation function used in first hidden layer	Weighted Myriad Filtering
3	Activation function used in second hidden layer	Hermitian multi-wavelet transform
4	Activation function used in third hidden layer	Schutz Feature Matching Coefficient
5	Learning rate	The value of the learning rate used in our work is 0.01
6	Number of epochs	The number of epochs is used in 20
7	Batch size	The value of the batch size used in our work is 16
8	Number of input images	100-1000 employed for experimentation

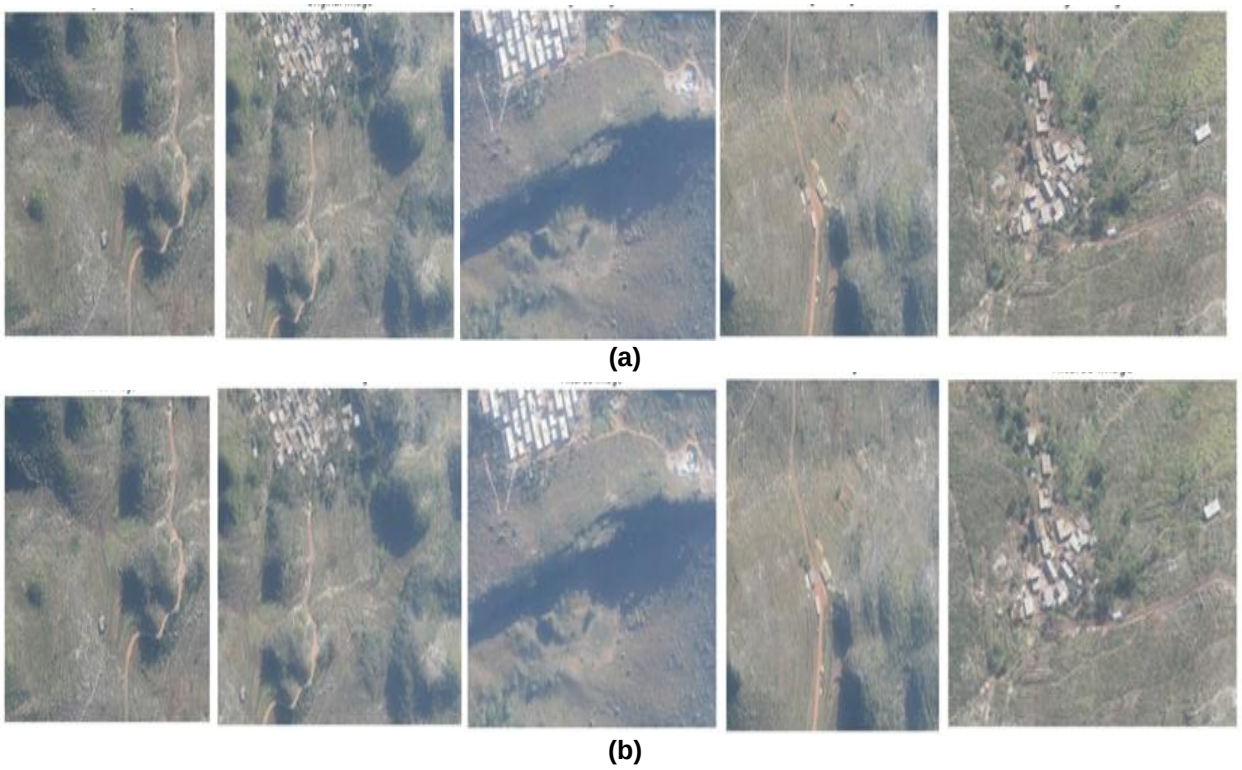

Figure 3 (a). Five sample input UAV images **(b)** Corresponding preprocessed images

Figure 3 shows five sample UAV images collected from the dataset. Each collected sample of UAV images, preprocessing is carried out by using the Weighted Myriad Filtering model to eradicate noisy images. Also, the preprocessed images are obtained, and it is given in Figure 3 (b). With preprocessed results, the features (texture, color, shape) are extracted with the aid of Hermitian wavelet transform in Figures 4, 5, 6, and 7. Finally, the building and non-building images are classified using the Schutz Feature Matching Coefficient in Figure 5.

Figure 4 illustrates the Hermitian wavelet transform of different decompositions of images of sub-band form five input images.

Figure 5 depicts the output results of texture feature extraction for different sample images.

Figures 6 and 7 show the results of shape and color feature extraction from sample images respectively.

Figure 8 demonstrates the final output of building detected images using the proposed WHWMELM technique.

Table 3 shows the comparison of the ground truth and predicted buildings. The ground truth and predicted building results are obtained from Figure 8. Here, accuracy is obtained as 90% and 88%. The overall quantitative performance analysis of diverse methods is examined during the following metrics.

Table 4 shows Description of performance parameters and Table 5 shows the performance analysis of PSNR for the proposed WHWMELM technique, existing BranTNet [1], multiscale fusion network [2] and WETUM [21] with and without denoise images. The PSNR is calculated for different classification methods. The above results confirm that the results of PSNR are improved in the WHWMELM technique than the conventional [1, 2] and [21]. Also, the results with denoising images are found to be better than those without denoising images. For instance, the PSNR using the proposed WHWMELM technique is obtained as 70.07 dB, BranTNet is obtained as 68.13 dB, the multiscale fusion network is obtained as 67.3 dB and WETUM [21] is obtained as 69.04 dB with the input of denoising image size 1.13 MB. Similarly, PSNR without denoising image (1.13 MB) is obtained as 68.13 dB, 66.54 dB, 65.85 dB and 67.71 dB using WHWMELM, BranTNet [1], and multiscale fusion network [2] and WETUM [21] is obtained as 69.04 dB.

The higher PSNR is obtained based on applying the Weighted Myriad filtering model significantly discovers the noisy pixels in input UAV images with a lesser amount of time. Then, the discovered noise pixels using the WHWMELM technique are taken away from the given image. With this, the proposed WHWMELM technique minimizes the mean square error. This supports for proposed WHWMELM technique to decrease the dissimilarity between original UAV images and preprocessed UAV images. Thus, WHWMELM enhances PSNR for building detection using denoising images by 10%, 7% and 2% as compared to BranTNet [1], Multiscale Fusion Network [2] and WETUM [21] respectively.

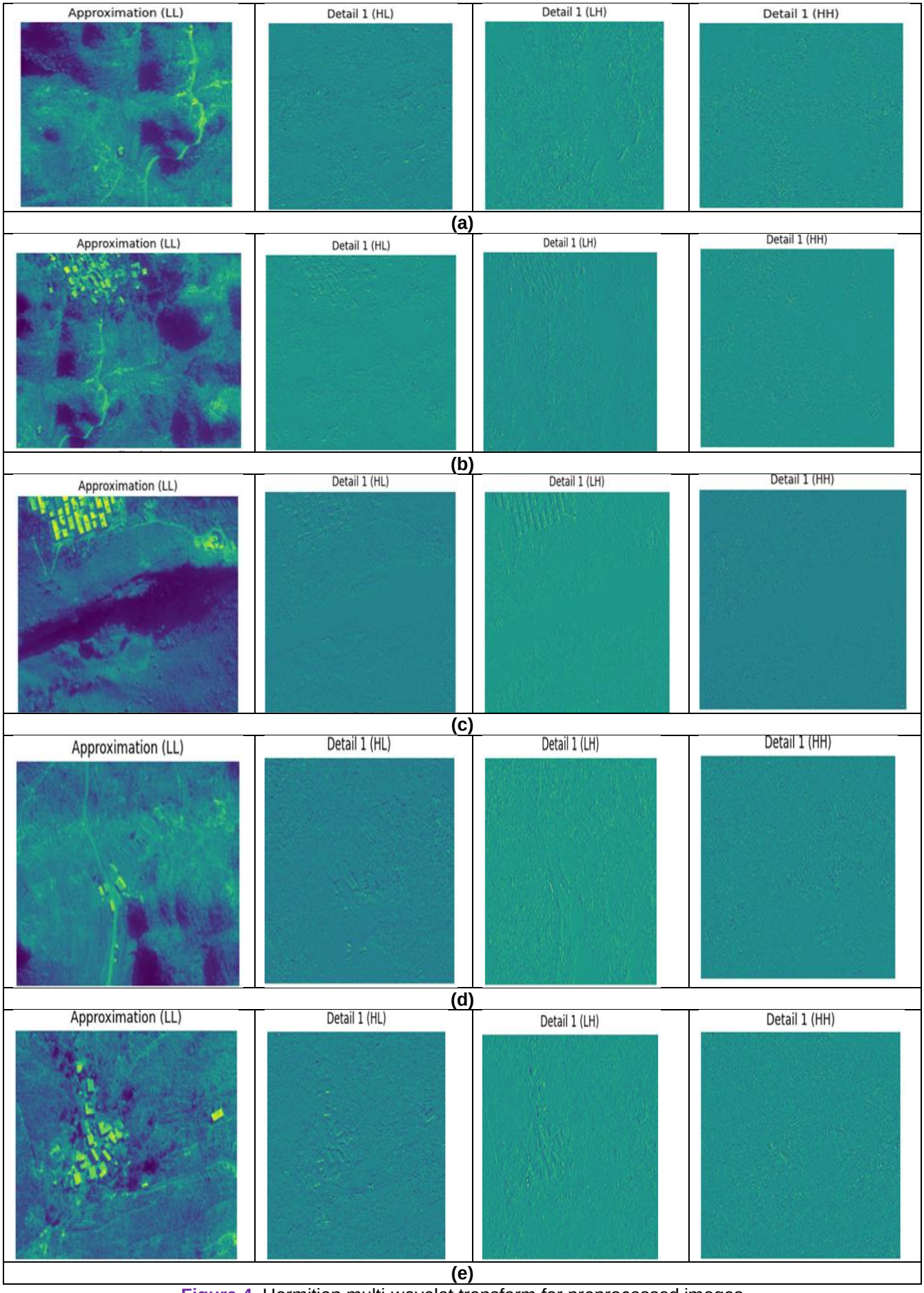

Figure 4. Hermitian multi-wavelet transform for preprocessed images

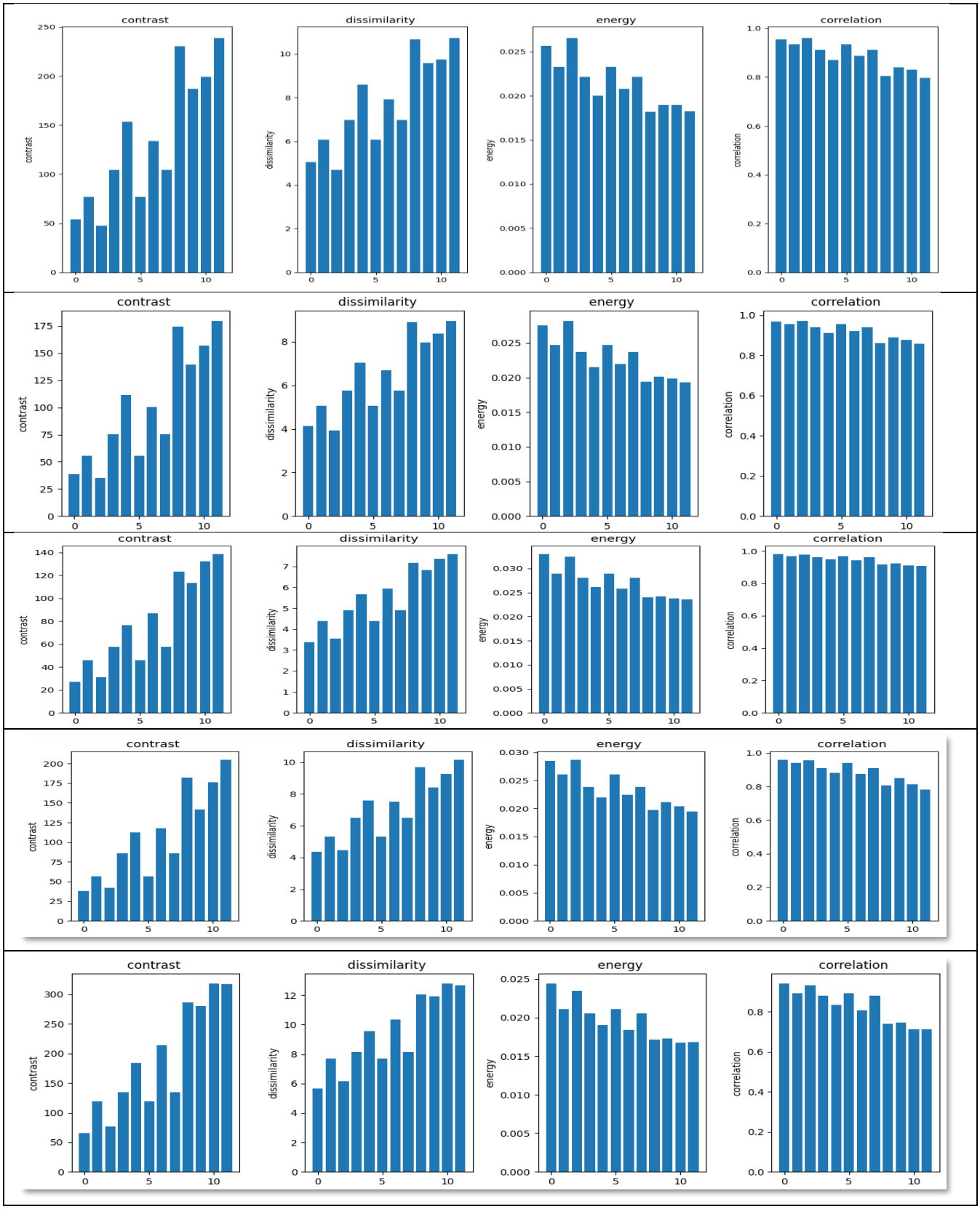

Figure 5. Texture feature extraction for sample images

Similarly, the WHWMELM technique improves PSNR up to 8%, 5% and 2% without denoising images as compared to BranTNet [1], Multiscale Fusion Network [2], and WETUM [21], respectively.

Figure 9 and 10 shows the classification accuracy of four different classification methods based on with and without denoising images collected from a given dataset for building detection. The results of classification accuracy are obtained for UAV 1000 images.

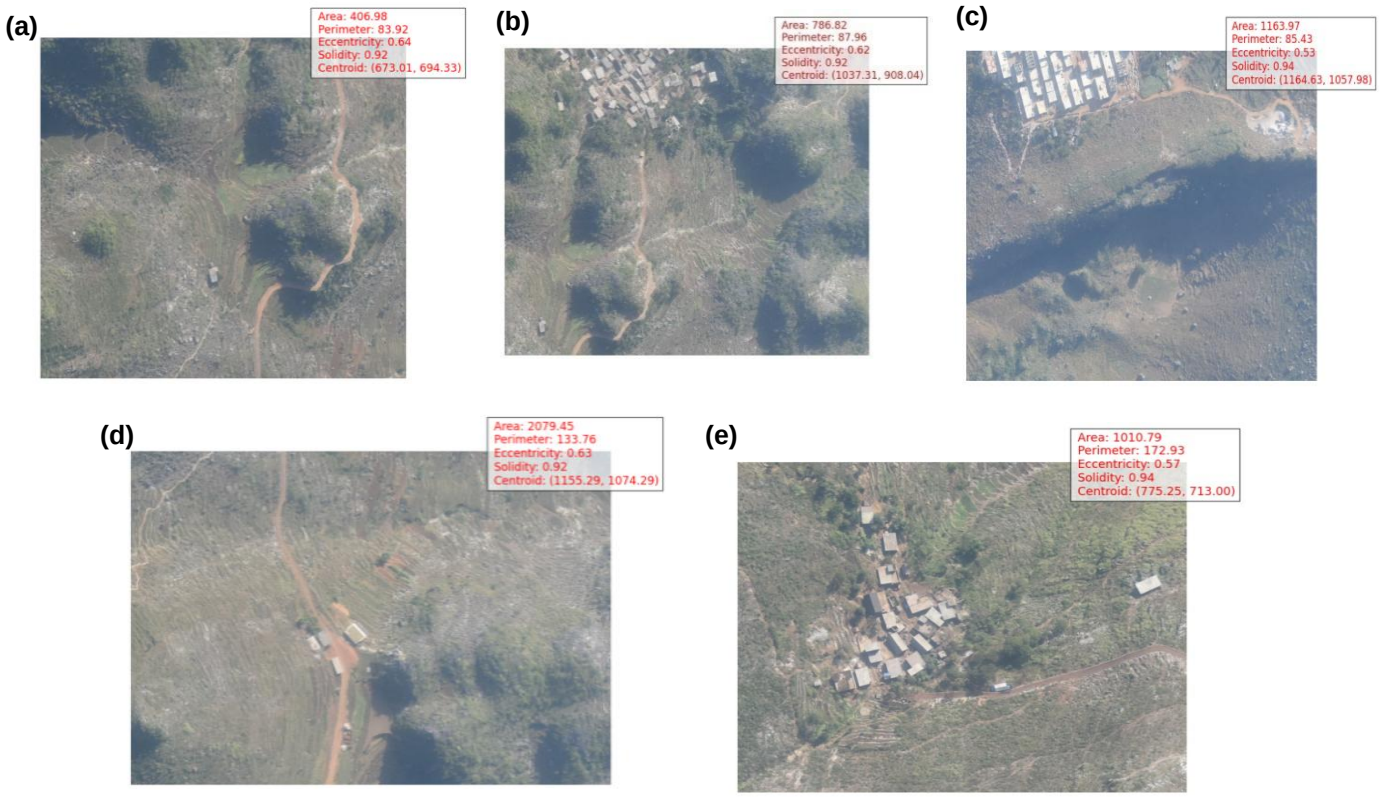

Figure 6. Shape feature extraction for sample images

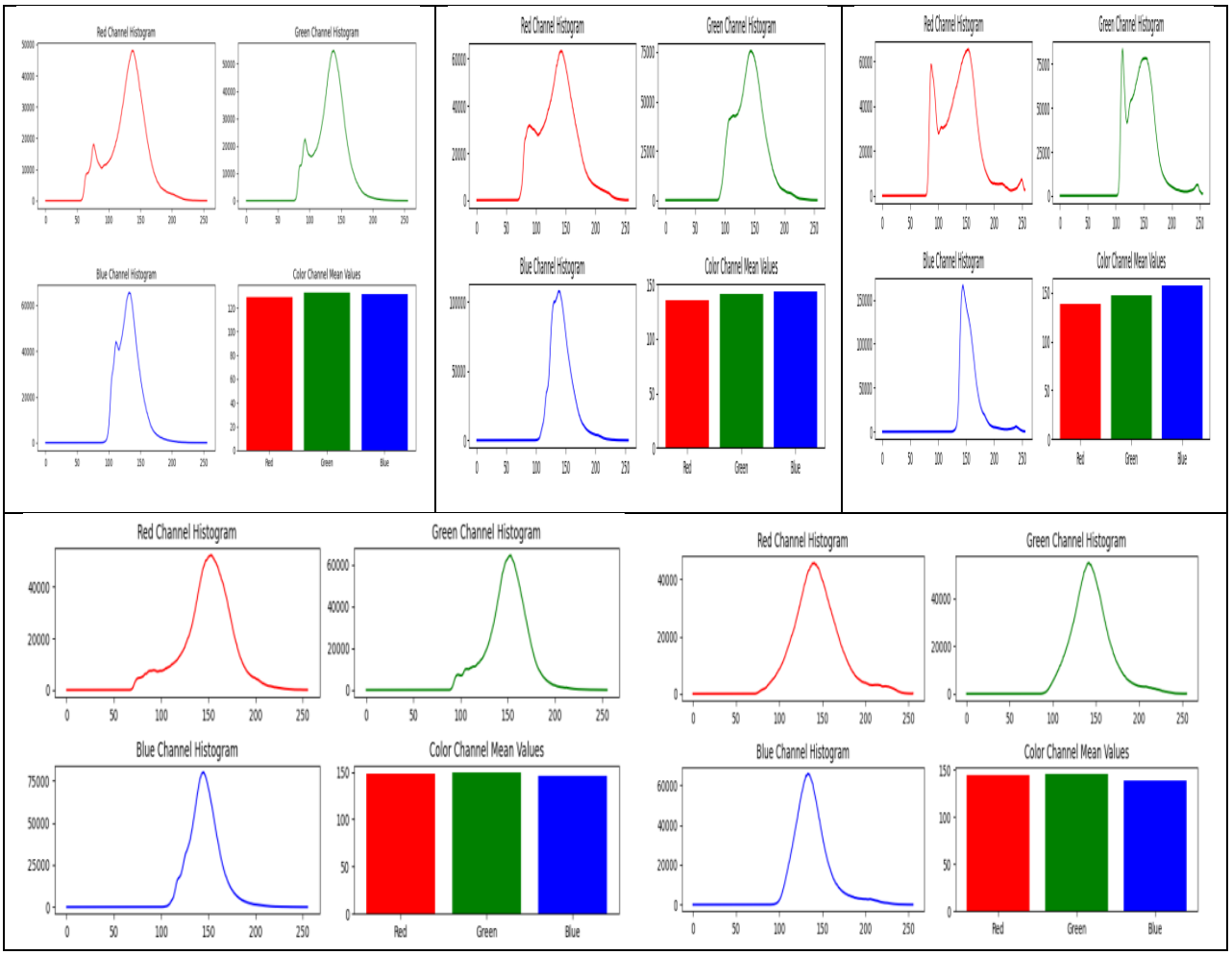

Figure 7. Color feature extraction for sample images

Figure 8. Output building detected images.

Table 3. Comparison of ground truth and predicted buildings

Number of ground truth building	Number of predicted building	Accuracy (%)
50	**45**	**90**
28	**25**	**88**

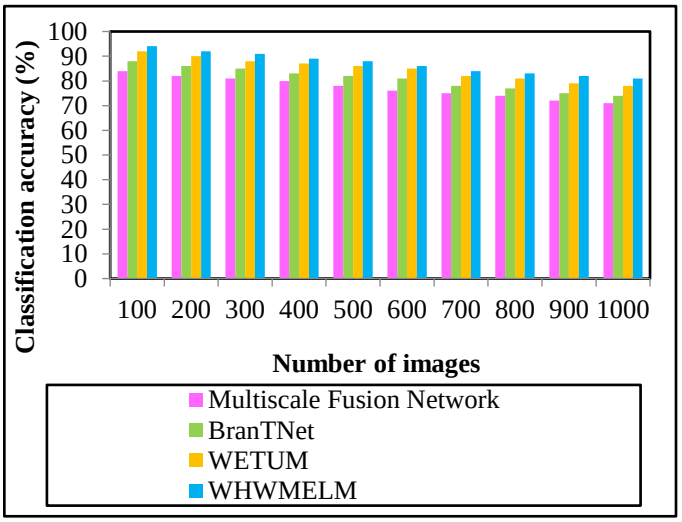

Figure 9. Classification accuracy with denoising images

The obtained outcomes of WHWMELM discussed with conventional BranTNet [1] and multiscale fusion network [2] and WETUM [21]. From the comparative graphs, the classification accuracy of images for building detection is improved using denoising images compared to without denoising images. In both the graphs, the accuracy value is

enhanced in the proposed WHWMELM technique than the existing [1, 2] and [21] methods.

On the contrary to other methods, the higher classification accuracy is achieved by applying the Schutz Feature Matching Coefficient in the WHWMELM technique for classifying the input UAV images. During the classification process, the similarity between extracted features from the images is computed. According to the similarity measure, images are classified into building objects or non-building objects. With this, the building objects in the UAV images are accurately identified for urban planning or development. The results of classification accuracy using the proposed WHWMELM technique with denoising images are increased by 13%, 8% and 3% compared to existing BranTNet [1], Multiscale Fusion Network [2] and WETUM [21] respectively. In addition, CA of WHWMELM without denoising images enhanced up to 12%, 9% and 3% compared to existing BranTNet [1] and Multiscale Fusion Network [2] and WETUM [21] respectively.

Table 6 describes the experimental results of training time with and without denoising images using the proposed WHWMELM technique, existing BranTNet [1], and multiscale fusion network [2] and WETUM [21]. Number of images is used as input and it is varied in the ranges of 100 to 1000.

Table 4. Description of performance parameters

Parameter	Symbol used	Description
Peak-Signal-to-Noise Ratio	$\mathrm{P_{snr}}$	It is computed depending on the mean square error which is estimated as the variation between the preprocessed image and original image. $$\boldsymbol{P_{snr} = 10 * \log_{10}\left[\frac{L^2}{M_{se}}\right]}$$ P_{snr} denotes a PSNR, L^2 refers to a maximum possible pixel range, M_{se} denotes a mean square error that is computed as the variation between preprocessing image size I_p and original input image size I_o.
Classification Accuracy	CA	Number of images correctly categorized for building detection as defined as CA for evaluation. $$\boldsymbol{CA = \frac{NI_{CC}}{T_I} * 100}$$ CA refers a classification accuracy, NI_{CC} refers number of images properly categorized.
Training time	TT	It computed as amount of time used by the algorithm to classify the images (building objects and non-building objects) for building detection. $$\boldsymbol{TT = I_n * Time\ (SI)}$$ I_n point outs the number of images and $Time\ (SI)$ refers a time taken for detecting building in a single image.
Precision	P	Precision is measured with number of true positives and false positives. $$P = \left(\frac{TP}{TP + FP}\right) * 100$$ P denotes precision, TP indicates true positive rate, and FP represents false positive rate.
Recall	R	Recall is calculated to determine the number of true positives and false negatives during the building detection. $$R = \left(\frac{TP}{TP + FN}\right)$$ R denotes recall, TP denotes true positive rate, and FN represents false negative rate.
F1-score	F	It is measured as the average value of both precisions and recall. $$F = \left[2 * \frac{P * R}{P + R}\right]$$
Memory Consumption	MC	**It** is referred to as the amount of memory consumed by algorithm for identifying the building. $$MO = \boldsymbol{I_n} * Mem\ (SI)$$ MO represents the memory consumption, and $Mem\ (SI)$' signifies the time taken for detecting building in single image. It is determined in terms of Mega byes (MB).

Table 5. Comparative results of PSNR

Image size (MB)	PSNR (dB)			
	Multiscale Fusion Network	BranTNet	WETUM	WHWMELM
With denoising images				
1.13	67.30	68.13	69.04	70.07
1.16	68.14	69.21	70.16	73.15
1.21	69.03	70.14	71.43	73.51
1.33	66.25	68.25	72.68	74.43
1.34	65.14	67.14	71.12	72.62
1.42	67.82	69.52	73.16	74.25
1.49	69.14	71.11	74.49	75.25
1.68	65.25	67.52	72.18	73.63
1.81	64.14	66.21	70.55	72.25
1.83	63.04	65.25	67.49	69.25
Without denoising images				
1.13	65.85	66.54	67.71	68.13
1.16	66.11	67.52	68.56	69.76
1.21	67.82	68.97	69.82	71.25
1.33	64.25	66.14	67.68	69.52
1.34	62.14	64.67	67.11	68.14
1.42	64.52	66.81	68.98	70.64
1.49	65.82	67.82	70.22	71.63
1.68	63.14	65.14	68.35	69.62
1.81	62.54	64.25	67.42	68.14
1.83	61.34	63	66.56	67.52

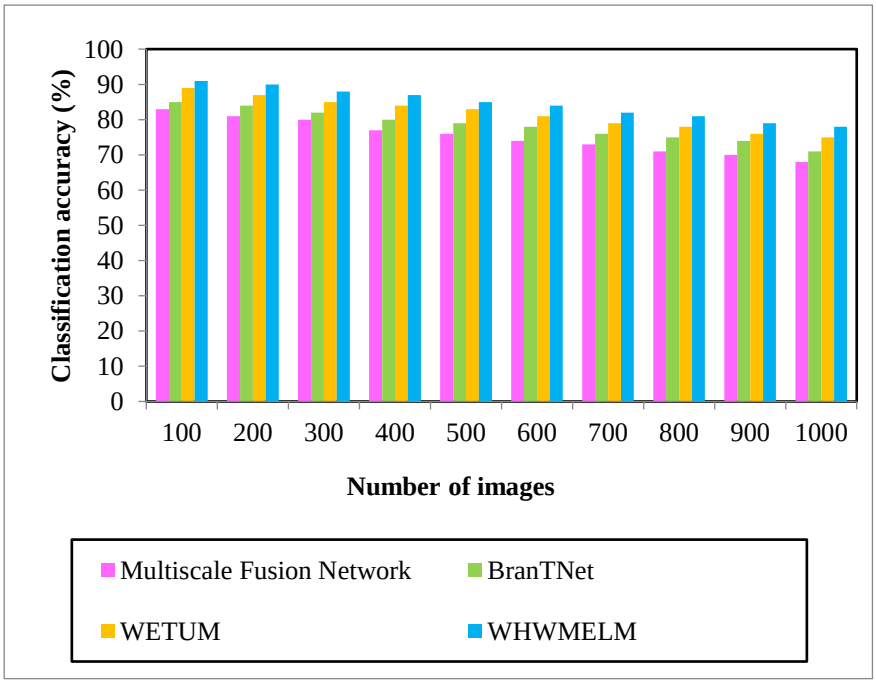

Figure 10. Classification accuracy without denoising images

Table 6. Comparative Results of Training Time

Number of images	Training time (ms)			
	Multiscale Fusion Network	**BranTNet**	**WETUM**	**WHWMELM**
With denoising images				
100	41	38	34	29
200	43	40	36	30
300	46	43	37	31
400	48	44	39	33
500	50	47	42	36
600	52	49	44	38
700	55	51	46	40
800	57	54	49	42
900	59	56	52	46
1000	60	58	54	48
Without denoising images				
100	46	43	39	32
200	48	46	41	35
300	49	47	42	36
400	52	49	44	40
500	55	52	46	42
600	56	53	48	45
700	59	55	49	47
800	60	57	52	49
900	62	59	54	51
1000	65	62	57	54

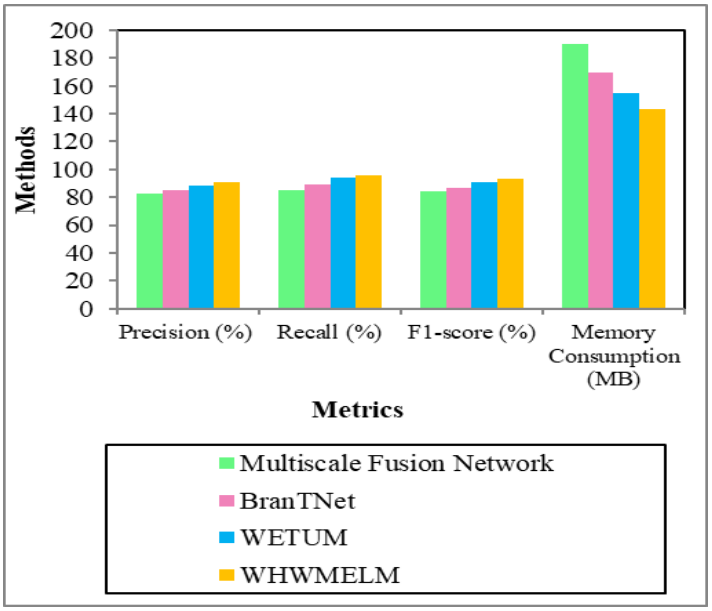

Figure 11. Performance of precision, recall, F1-score and memory consumption with denoising images

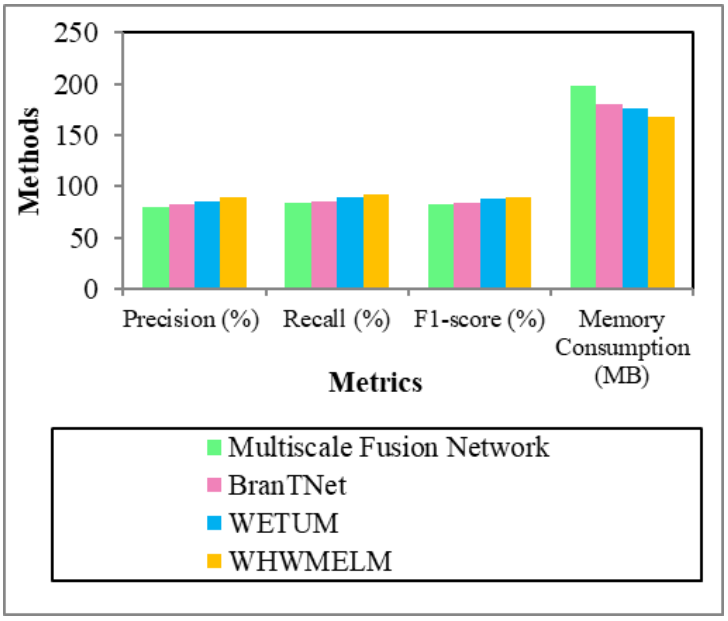

Figure 12. Performance of precision, recall, F1-score, and memory consumption with denoising images

From the table 6, the training time of the proposed WHWMELM technique is found to be lower for building detection than the conventional methods. As compared to without denoising images, training time is minimized with denoising images. With the input of 100 images, the classification time is measured to be 29ms, 38ms, 41ms and 34ms using the WHWMELM technique, existing BranTNet [1], multiscale fusion network [2] and WETUM [21]. By considering the input without denoising images, the training time is obtained as 32ms, 43ms, and 46ms and 39ms for proposed and existing methods. The lesser time consumption is attained by performing the most relevant feature extraction process in the WHWMELM technique. Images to different regions decomposed by hermitian multi-wavelet transform. From the decomposed images, building footprints detected to extract features. These extracted features are classified with lower time through the coefficient matching process.

This helps to reduce the time needed for classification in the WHWMELM technique than the other methods. The results of training time for denoising images are minimized using the proposed WHWMELM technique by 27% and 23% and 14% as compared to the existing BranTNet [1] and multiscale fusion network [2] and WETUM [21]. Training time output of the proposed WHWMELM technique without denoising the image is reduced by 22%, 18% and 9% to the existing [1] [2] and [21] methods.

Figures 11 and 12 show the performance results of precision, recall, F1-score, and memory consumption with and without denoising images using the proposed WHWMELM technique, existing BranTNet [1], and multiscale fusion network [2], and WETUM [21].Proposed WHWMELM technique outperforms existing methods in terms of achieving higher precision, recall, F1-score and lesser memory consumption. This improvement is achieved by employing the Schutz Feature Matching Coefficient to determine whether images as building objects or non-building objects. The results of precision for denoising images are improved using the proposed WHWMELM technique by 7%, 10% and 3% as compared to the existing BranTNet [1] and multiscale fusion network [2], and WETUM [21]. Precision output of the proposed WHWMELM technique without denoising the image is enhanced by 9%, 11% and 3% than the existing [1, 2] and [21] methods.

The results of recall for denoising images are improved using the proposed WHWMELM technique by 9%, 13% and 2% as compared to the existing BranTNet [1] and multiscale fusion network [2], and WETUM [21]. Recall output of the proposed WHWMELM technique without denoising the image is improved by 7%, 11% and 2% than the existing [1, 2] and [21] methods. The results of the F1-score for denoising images are improved using the proposed WHWMELM technique by 7%, 11% and 2% as compared to the existing BranTNet [1] and multiscale fusion network [2], and WETUM [21]. F1-score output of the proposed WHWMELM technique without denoising the image is enhanced by 7%, 10% and 2% than the existing [1, 2] and [21] methods. Also, the space complexity for denoising images is decreased using the proposed WHWMELM technique by 16%, 25% and 8% as compared to the existing BranTNet [1] and multiscale fusion network [2], and WETUM [21]. Space complexity output of the proposed WHWMELM technique without denoising the image is reduced by 7%, 15% and 5% to the existing [1, 2] and [21] methods.

6. Conclusion

A novel proposed a detection procedure named WHWMELM is developed for building detection using UAV images with higher accuracy and less time. Initially, the Weighted Myriad Filtering model-based

preprocessing is carried out to eliminate the noisy pixels in the images to increase the image quality. After that, texture, color, and shape features were extracted with the Hermitian multi-wavelet transform. With the extracted features, the time consumed for building detection is reduced. The designed technique carried out the classification of building images using the Schutz Feature Matching Coefficient. The similarity between extracted features and testing features is computed to classify the images into building or non-building objects. From that, the UAV images are classified with higher accuracy and smaller errors. Experimental evaluation of the WHWMELM technique is carried out with diverse factors with respect to the UAV images. The result demonstrates that the proposed WHWMELM technique improved the performance of building detection in terms of higher classification accuracy of 8%, PSNR of 10%, precision of 7%, recall of 8% and F1-score of 7% and lesser time of 21% and memory consumption of 16% with denoising images as compared to existing methods. Outcomes of the WHWMELM technique are to provide maximum PSNR by 5%, accuracy by 8%, precision by 8%, recall by 6% and F1-score by 6% and minimum time 16% and memory consumption by 9% without denoising images than the conventional methods. The limitation of the proposed technique is that considers only one dataset. In future work, the proposed technique is extended to consider more datasets for enhancing the building detection performance.

References

[1] H. Wang, Z. Yao, T. Li, Z. Ying, X. Wu, S. Hao, M. Liu, Z. Wang, T. Gu, Enhanced open biomass burning detection: The BranTNet approach using UAV aerial imagery and deep learning for environmental protection and health preservation. Ecological Indicators, 154, (2023) 110788. https://doi.org/10.1016/j.ecolind.2023.110788

[2] C. Zheng, B. Peng, B. Chen, M. Liu, W. Yu, Y. He, D. Ren, Multiscale fusion network for rural newly constructed building detection in unmanned aerial vehicle imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, (2022) 9160-9173. https://doi.org/10.1109/JSTARS.2022.3209682

[3] J. Liu, H. Huang, H. Sun, Z. Wu, R. Luo, LRAD-Net: An improved lightweight network for building extraction from remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, (2022) 675-687. https://doi.org/10.1109/JSTARS.2022.3229460

[4] R.J. Batista, d. Silva, R. Rodrigues, Dias, J. Dantas, Escarpinati, M. Cunha, Backes, A. Ricardo, Automated detection of sugarcane crop lines from UAV images using deep learning. Information Processing in Agriculture, 11(3), (2024) 385-396. https://doi.org/10.1016/j.inpa.2023.04.001

[5] L.M.A. dos Santos, V.A.G. Zanoni, E. Bedin, H. Pistori, Deep learning applied to equipment detection on flat roofs in images captured by UAV. Case Studies in Construction Materials, 18, (2023) e01917. https://doi.org/10.1016/j.cscm.2023.e01917

[6] P. Ong, K.S. Teo, C.K. Sia, UAV-based weed detection in Chinese cabbage using deep learning. Smart Agricultural Technology, 4, (2023) 100181. https://doi.org/10.1016/j.atech.2023.100181

[7] F. Moradi, F.D. Javan, F. Samadzadegan, Potential evaluation of visible-thermal UAV image fusion for individual tree detection based on convolutional neural network. International Journal of Applied Earth Observation and Geoinformation, 113, (2022) 103011. https://doi.org/10.1016/j.jag.2022.103011

[8] C. Jiang, H. Ren, X. Ye, J. Zhu, H. Zeng, Y. Nan, M. Sun, X. Ren, H. Huo, Object detection from UAV thermal infrared images and videos using YOLO models. International Journal of Applied Earth Observation and Geoinformation, 112, (2022) 102912. https://doi.org/10.1016/j.jag.2022.102912

[9] S. Liu, X. Jin, Y. Bai, W. Wu, N. Cui, M. Cheng, Yadong Liu, L. Meng, X. Jia, C. Nie, D. Yin, UAV multispectral images for accurate estimation of the maize LAI considering the effect of soil background. International Journal of Applied Earth Observation and Geoinformation, 121, (2023) 103383. https://doi.org/10.1016/j.jag.2023.103383

[10] P.K. Yadav, J.A. Thomasson, R. Hardin, S.W. Searcy, U. Braga-Neto, S.C. Popescu, D.E. Martin, R. Rodriguez, K. Meza, J. Enciso, J.S. Diaz, T. Wang, Detecting volunteer cotton plants in a corn field with deep learning on UAV remote-sensing imagery. Computers and Electronics in Agriculture, 204, (2023) 107551. https://doi.org/10.1016/j.compag.2022.107551

[11] B. Fu, P. Zuo, M. Liu, G. Lan, H. He, Z. Lao, Y. Zhang, D. Fan, E. Gao, Classifying vegetation communities karst wetland synergistic use of image fusion and object-based machine learning algorithm with Jilin-1 and UAV multispectral images. Ecological Indicators, 140, (2022) 108989. https://doi.org/10.1016/j.ecolind.2022.108989

[12] C.J. Lee, M.D. Yang, H.H. Tseng, Y.C. Hsu, Y. Sung, W.L. Chen, Single-plant broccoli growth monitoring using deep learning with UAV imagery. Computers and Electronics in Agriculture, 207, (2023) 107739. https://doi.org/10.1016/j.compag.2023.107739

[13] Z. Ye, K. Yang, Y. Lin, S. Guo, Y. Sun, X. Chen, R. Lai, H. Zhang, A comparison between Pixel-

based deep learning and Object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV Visible-light images. Computers and Electronics in Agriculture, 209, (2023) 107822. https://doi.org/10.1016/j.compag.2023.107822

[14] S.K. Srivastava, K.P. Seng, L.M. Ang, Anibal 'Nahuel' A. Pachas, T. Lewis, Drone-Based Environmental Monitoring and Image Processing Approaches for Resource Estimates of Private Native Forest, Sensors, MDPI, 22(20), (2022) 7872. https://doi.org/10.3390/s22207872

[15] M.A. Ahmed, J. Aloufi, S. Alnatheer, Satin bowerbird optimization with convolutional LSTM for food crop classification on UAV imagery. IEEE Access, 11, (2023) 41075-41083. https://doi.org/10.1109/ACCESS.2023.3269806

[16] C. Kyrkou, T. Theocharides, EmergencyNet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, (2020) 1687-1699. https://doi.org/10.1109/JSTARS.2020.2969809

[17] M. Liu, H. Deng, W. Dong, Identification of mangrove invasive plant derris trifoliate using UAV images and deep learning algorithms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, (2022) 10017-10026. https://doi.org/10.1109/JSTARS.2022.3223227

[18] M. Ariza-Sentís, H. Baja, S. Vélez, J. Valente, Object detection and tracking on UAV RGB videos for early extraction of grape phenotypic traits. Computers and Electronics in Agriculture, 211, (2023) 108051. https://doi.org/10.1016/j.compag.2023.108051

[19] W. Xing, Z. Cui, J. Qi, HRCTNet: a hybrid network with high-resolution representation for object detection in UAV image. Complex & Intelligent Systems, 9(6), (2023) 6437-6457. https://doi.org/10.1007/s40747-023-01076-6

[20] C. Song., F. Zhang, Y.A.N.G. Chen, Z.H.O.U. Hang, J.X. Zhang, (2023). Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. Journal of Integrative Agriculture, 22(6), 1671-1683. https://doi.org/10.1016/j.jia.2022.09.021

[21] E. Khankeshizadeh, A. Mohammadzadeh, H. Arefi, A. Mohsenifar, S. Pirasteh, E. Fan, H. Li, J. Li, A novel weighted ensemble transferred U-Net based model (WETUM) for postearthquake building damage assessment from UAV data: A comparison of deep learning-and machine learning-based approaches. IEEE Transactions on Geoscience and Remote Sensing, 62, (2024) 1-17. https://doi.org/10.1109/TGRS.2024.3354737

[22] T. Hou, J. Li, Application of mask R-CNN for building detection in UAV remote sensing images. Heliyon, 10(19), (2024). https://doi.org/10.1016/j.heliyon.2024.e38141

[23] Q. Fan, Y. Li, M. Deveci, K. Zhong, S. Kadry, LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle. Information Sciences, 686, (2025) 121366.

[24] C. Tan, Chen, T., Liu, J., Deng, X., Wang, H., & Ma, J. (2024). Building extraction from Unmanned Aerial Vehicle (UAV) data in a landslide-affected scattered mountainous area based on res-unet. Sustainability, 16(22), 9791. https://doi.org/10.3390/su16229791

[25] W. Alsabhan, T. Alotaiby, B. Dudin, Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. Computational Intelligence and Neuroscience, 2022(1), (2022) 4831223. https://doi.org/10.1155/2022/4831223

Authors Contribution Statement

A. Franklin Alex Joseph: Conceptualization, Methodology, Literature Review, writing original draft. A. Manikandan: Data Curation. Umi Salman Basha: Data Curation. Shanmugapriya Velmurugan: Formal analysis, Validation. Rajesh Natarajan: Formal analysis, Validation. S. Subash Chandra Bose: Writing-Original Draft, Supervision. All Authors Read and Approved Final version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

UAV Small Object Detection Dataset is publicly available within the article. The sample images taken from https://www.kaggle.com/datasets/sovitrath/uav-small-object-detection-dataset

Has this article screened for similarity?

Yes