Integrating Machine Learning and OBIA for Vegetation Classification in Archived Grayscale Aerial Imagery

1st Kumpee Teeravech
Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University,
Chanthaburi/Thailand
kumpee.t@rbru.ac.th

2nd Phummipat Ounban
Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University,
Chanthaburi/Thailand
phummipat.o@rbru.ac.th

3rd Wira Srimala
Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University,
Chanthaburi/Thailand
wira.s@rbru.ac.th

4th Taweesak Samma
Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University,
Chanthaburi/Thailand
taweesak.s@rbru.ac.th

5th Kraisri Sawangsri
Department of Special Investigation,
Ministry of Justice,
Nonthaburi/Thailand
kraisri_sa@dsi.go.th

6th Kittakorn Viriyasatr
Defence Technology Institute,
Nonthaburi/Thailand
kittakorn.v@dti.or.th

7th Warakorn Luangluewut
Defence Technology Institute,
Nonthaburi/Thailand
warakorn.l@dti.or.th

8th Chamnan Kumsap
Defence Technology Institute,
Nonthaburi/Thailand
chamnan.k@dti.or.th

***Abstract*— This study explores the use of machine learning models to classify water, vegetation, and non-vegetation land cover types in archived grayscale aerial imagery. The input images are segmented using a superpixel algorithm, and the resulting segments are mapped to expert-provided reference data. The region-based and patch-based approaches are evaluated using artificial neural networks and convolutional neural networks, respectively. The region-based method achieves an average accuracy of 0.83, while the patch-based method reaches 0.79. Although the patch-based method shows slightly lower overall accuracy, it significantly improves recall rates, particularly for the water and non-vegetation classes.**

Keywords*— *land use/land cover classification, object-based image analysis, gray-level co-occurrence matrix

I. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) technologies now play a significant role in processing large-scale data, including the analysis of remotely sensed imagery. These technologies are increasingly being used as key evidence in cases involving natural resources and environmental crimes. In Thailand, the Department of Special Investigation (DSI), part of the Ministry of Justice, is responsible for investigating and prosecuting cases of natural resource destruction or exploitation, frequently using historical grayscale aerial imagery as crucial evidence. For decades, the analysis of these archived imagery depended heavily on the expertise of highly specialized professionals, whose limited availability in Thailand frequently caused delays and challenges. Thus, the integration of automated techiniques is expected to enhance the efficiency of remotedly sense image analysis.

In remote sensing, Object-Based Image Analysis (OBIA) analyzes both the spectral and spatial properties of pixel groups, unlike traditional pixel-wise methods. The overall workflow of OBIA in remote sensing typically begins with segmenting the input images into meaningful objects or pixel groups based on their spectral and spatial properties. These segments are then classified using various techniques to identify specific land use/land cover types [1][2][3]. The Gray-Level Co-Occurrence Matrix (GLCM) is commonly used for texture analysis in OBIA, with the ability to improve classification accuracy in applications like land cover mapping and environmental monitoring [1]. In [4], GLCM is used for landslide detection in multispectral imagery, while [5] studied its applications with optical and SAR (synthetic aperture radar) imagery. In [6], the authors combined GLCM, convolutional neural networks (CNN), and other machine learning techniques to analyze very high resolution (VHR) satellite images. Futhermore, tree species in aerial imagery can also be anaylzed and classified as demonstrated in [7]. Therefore, the primary objective of this study is to explore automated methods that integrate machine learning and OBIA techniques for classifying of vegetation land cover type in historical or archived grayscale aerial orthophotos.

II. METHODOLOGY

Figure 1 illustrates the overall methodology of this study, where the input data consists of serveral individual scenes of archived grayscale aerial orthophotos provided by multiple sources. These orthophotos are firstly merged and then segmented using a superpixel algorithm. After that, the ground truth maps from the experts, along with the information from the segmented images are used to preparethe region of interests (ROI). After that image segmentation steps, our experiment is divided into two approaches: a region-based

979-8-3503-6630-3/24/$31.00 ©2024 IEEE

method and a patch-based method. In the region-based approach, various properties of the segments are extracted, as is commonly done in OBIA methods. In contrast, in the patch-based approach, image patches centered around the centroid of each segment are extracted. Since the output of the region-based method is a tabular dataset, an artificial neural network (ANN) is employed for classification. On the other hand, the patch-based method deals with image data, so a convolutional neural network is used for classification. The outputs of both methods are evaluated against ground truth maps provided by external experts before serving to the target users via web services. Detailed descriptions of each step are provided in the following sections.

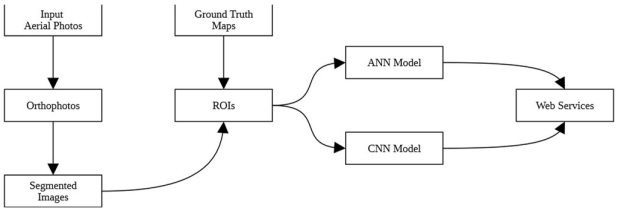


Figure 1. Overall workflow of the our study.

A. Study Area

The study area is located in Saiyok, a district in Kanchanaburi province in western Thailand. The total area covers 25.08 square kilometers. Figure 2 provides an overview map of the study area.

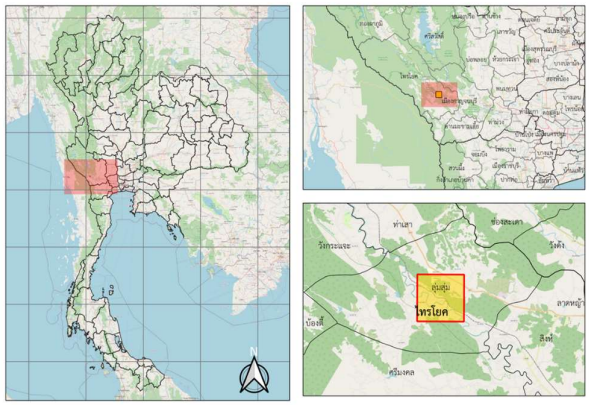


Figure 2. Study area.

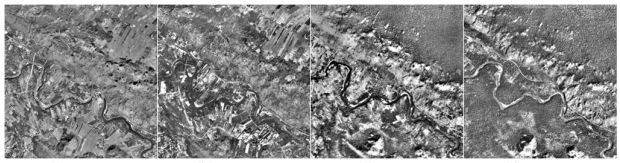
B. Datasets Preparation

The primary data sources for this study are aerial orhophotos taken between 1966 and 2003 by various projects, including the Department of Lands (DOL), the National Imagery and Mapping Agency (NIMA), the Vertical Aerial Photograph Project (VAP), and the World Wide Survey (WWS) projects. Please note that one of the datasets originally consists of RGB orthophotos, but has been converted to grayscale to align with the objective of this study. Table 1 summarizes the characteristics of each dataset.

TABLE 1. DATASETS.

Name	Extended Year	Resolution (m)	No. of Images
DOL	1986-2001	0.75	16
NIMA	1996-2000	0.75	16
VAP	1966-1970	0.50	16
WWS	1952-1956	0.50	16

For each dataset, all individual aerial orhophotos are merged to create a single orthophoto. The resulting orthophoto is then subset to match the target study area and resampled to the target resolution of 0.75 meters per pixel, as shown in Figure 3.



(a) DOL (b) NIMA (c) VAP (d) WWS

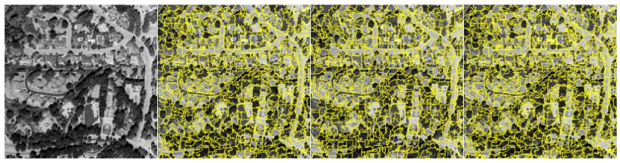
Figure 3. Comparison of the datasets.

Figure 4 represents the study area, which is outlined by a red rectangle, while the boundary of each individual input aerial orhophotos are illustrated by yellow rectangles.

Figure 4. Study area and dataset availability.

C. Image Segmentation

To extract meaningful spectral and spatial characteristics, the orthophoto is segmented into smaller regions using superpixel algorithm. This study use OpenCV as the primary digital image processing package, which offers several superpixels algorithms in its extended image processing library, including the Linear Spectral Clustering (LSC) [8], Seeds [9] and the Simple Linear Iterative Clustering (SLIC) [8]. In the case of SLIC, Several optimizations are available, such as SLICO [8], and MSLIC [10]. Based on our preliminary results, as shown in Figure 5, we selected the SLIC algorithm as it provided the most satisfactory output, i.e., maintaining the approximate shapes of the objects better than the others. The parameters that control the output of the SLIC algorithm are the region size which determines the average superpixel size, measured in pixels, and the ruler that enforces the smoothness of superpixels.



(a) Input (b) SLIC (c) SLICO (d) MSLIC

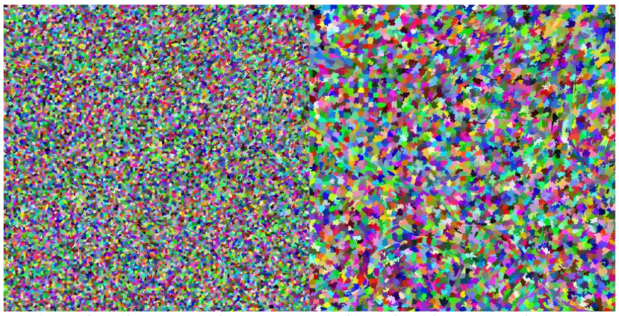
Figure 5. Examples of segmentations by SLIC, SLICO, and MSLIC.

To select appropriate values for these parameters, we conducted preliminary experiments by varying the region size and ruler values within the ranges [10, 20, …, 100, 150] and [5, 10, …, 75, 100], respectively. Table 2 provides examples of the number of extracted segments for different combinations of region size and ruler values.

TABLE 2. TESTING PARAMETERS FOR IMAGE SEGMENTATION.

Region Size\Ruler	60	80	100	150
10	25246	13741	8435	3544
20	27340	14921	9243	3905
30	28085	15512	9727	4158
40	28273	15736	9907	4310
50	28269	15813	9972	4368
75	28104	15860	10081	4478
100	28072	15870	10096	4484

Figure 6 visualizes examples of this parameters selection process. Based on our observations, a region size of 55 and a ruler value of 35 were selected, as they give the best balance between preserving the shape of the original objects and maintaining segment compactness.



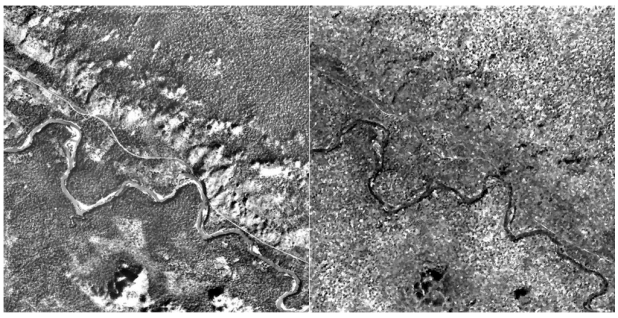
(a) region size = 40, ruler = 30 (b) region size = 80, ruler = 30

(c) region size = 150, ruler = 20 (d) region size = 150, ruler = 100

Figure 6. Example results of varing SLIC parameters.

D. Region-based Feature Extractions

The common properties to be extracted from the segments are intensity, shape, and texture. The intensity properties, such as brightness, mean, and standard deviation, represent the light reflectance of objects within the region. Shape properties, including area, perimeter, and length-to-width ratio, describe the geometric characteristics of the segments. Figure 7 visualizes examples of mean and standard deviation extracted from the WWS dataset.



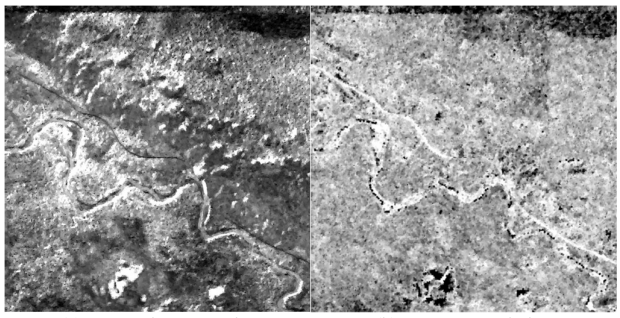
(a) Mean (b) Standard deviation

Figure 7. Example visualizationf of the extracted intensity properties.

The final group of properties is texture, typically extracted using the GLCM. It captures texture by analyzing how pixel intensity values relate to their neighboring pixels. Various features can be derived from the GLCM, such as contrast, dissimilarity, and homogeneity. In this study, the GLCM was calculated with a window size of 3x3 pixel and angles of [0, 90, 180, 270] degrees. Based on our preliminary tests, a complete list of the features used in this study is provided in Table 3. Figure 8 visualizes examples of GLCM homogeneity and GLCM correlation extracted from the WWS dataset.

TABLE 3. FEATURES USED FOR THE ANALYSIS.

Attributes	Properties
Intensity	Mean Standard deviation
Shape	Length/width ratio Perimeter Solidity
Texture	GLCM texture GLCM contrast GLCM homogeneity GLCM energy GLCM correlation GLCM dissimilarity GLCM ASM

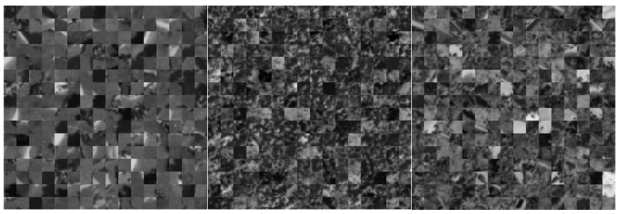


(a) GLCM homogeneity (b) GLCM correlation

Figure 8. Example visualizationf of the extracted GLCM properties.

E. Patch-based Feature Extraction

In the patch-based feature extraction method, instead of using GLCM to extract segment properties, a 64x64 pixel image patch centered around the centroid of each segment is extracted and treated as a standard image classification problem. Figure 9 shows examples of extracted image patches for water, vegetation, and non-vegetation classes.

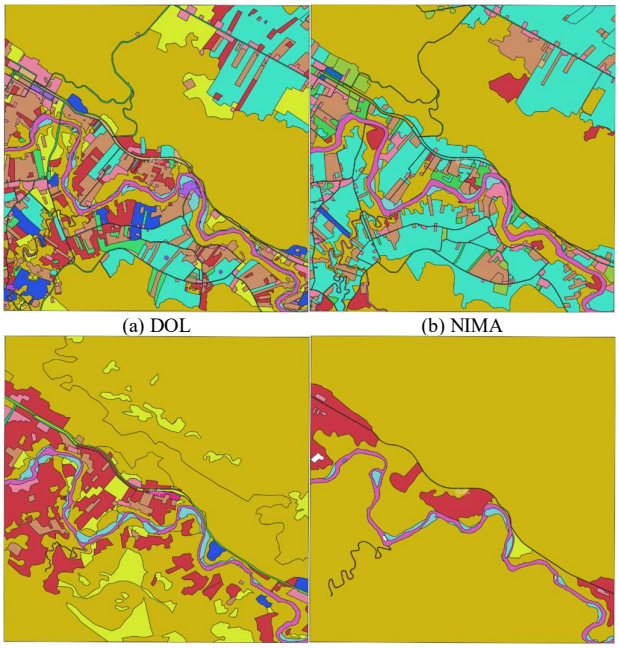


(a) Water (b) Vegetation (c) Non-vegetaion

Figure 9. Example of the image patches extracted from WWS dataset.

F. Training Data Preparation

The next step is to map the extracted features to the desired output classes. The ground truth maps for each dataset, originally provided by external experts, are shown in Figure 10.



(a) DOL (b) NIMA

(c) VAP (d) WWS

Figure 10. Ground truth maps.

TABLE 4. MAPPING BETWEEN SOURCE LAND USE CLASSES AND THE TARGET LAND COVER CLASSES.

Source-ID	Target-ID	Source-ID	Target-ID
A0	V	U2	N
A1	V	U203	N
A2	V	U3	N
A3	V	U402	N
A4	V	U405	N
A5	V	U406	N
F	V	U6	N
M	N	U601	N
M1	V	U602	N
M402	N	W1	W
M405	N	W2	W
M406	N	W202	W

However, these reference maps represent land use classes, some of which are too detailed and beyond the scope of this study. Additionally, not all classes are present across every datasets. Therefore, we reclassified the input land use classes into our target land cover classes, as shown in Table 4. Specifically, all original classes containing or covered by vegetation, such as forests and agricultural areas, were grouped into the vegetation class, denoted as V. Water bodies and areas covered by water were reclassified as water, denoted as W. All remaining land use types were assigned to the non-vegetation category, denoted as N.

After mapping features to the output classes, 80% of the segments are randomly selected as the training dataset, while the remaining 20% are used as the testing dataset. Finally, the feature values for each segment are exported in CSV file format.

G. AI Modeling

In this study, the ANN and CNN models for vegetation vs non-vegetation classification are developed, trained, and evaluated using SONY Neural Network Console (NNC) software. SONY NNC is a tool from SONY designed for creating, training, and deploying neural network models without the need for coding knowledge. Its drag-and-drop graphical user interface makes model design easier compared to traditional programming.

In this study, we built the models from scratch while aiming to minimize their size and complexity, as we plan to deploy them in both web-based and microcontroller-based applications in the future. To achieve this, the convolutional and fully connected layers were carefully configured, and the accuracy of each trial model was evaluated. Figure 11 represents the UI of SONY NNC with an example CNN model that takes a single-channel 28x28 image patch as input and classifies it into three classes.

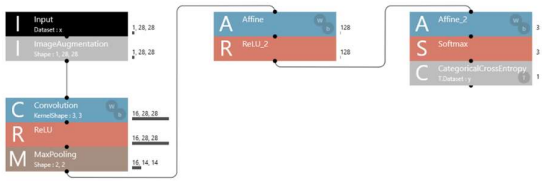


Figure 11. Example of designing a CNN model in SONY NNC.

III. RESULTS

The source aerial images are processed using QGIS 3.28.11. Python 3.9 is used as the programming language. Image segmentation with the SLIC algorithm, along with most other image processing tasks, was performed using the extended image processing module of OpenCV 4.5.2. The GLCM and related properties were extracted using the scikit-image package. For machine learning tasks, SONY NNC 3.1 with GPU acceleration was used. We used the Adam optimizer with the following parameters: epoches=1000, batch size=128, alpha=0.001, beta1=0.9, beta2=0.999, and epsilon=1e-8. All experiments were conducted on a PC running Windows 11, equipped with an Intel i7-12700K 3.6GHz CPU, 64GB DDR5 RAM, 2GB NVMe storage, and a NVIDIA GeForce RTX 3060 display card. The details of the experimental results are as follows.

A. Region-based Method

Figure 12 shows the ANN model used for the DOL dataset, which generated the results shown in Table 5. It consists of 12 dense layers with the number of neurons ranging from 1024 to 4096. Table 6 to Table 8 present the classification results of this ANN model for the remaining datasets where the accuracy rates for the DOL, NIMA, VAP, and WWS datasets are 0.80, 0.85, 0.77, and 0.90, respectively, resulting in an average accuracy of 0.83. The vegetation class consistently has the highest recall rate, exceeding 0.90 in most cases, while the others have recall rates below 0.60.

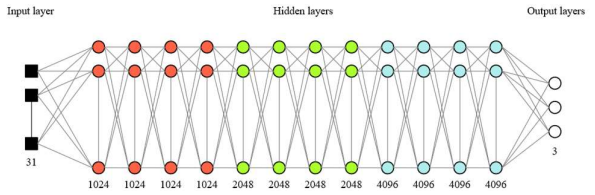


Figure 12. The resultant ANN model.

TABLE 5. CONFUSION MATRIX OF THE DOL DATASET (ANN).

	W	V	N	Recall
W	148	42	23	0.6948
V	5	4579	416	0.9158
M	8	877	913	0.5077
Precision	0.9192	0.8328	0.6752	
F-Measures	0.7914	0.8723	0.5795	

TABLE 6. CONFUSION MATRIX OF THE NIMA DATASET (ANN).

	W	V	N	Recall
W	95	171	12	0.3417
V	14	4800	186	0.9600
M	2	486	403	0.4523
Precision	0.8558	0.8796	0.6705	
F-Measures	0.4883	0.918	0.5401	

TABLE 7. CONFUSION MATRIX OF THE VAP DATASET (ANN)

	W	V	N	Recall
W	51	90	6	0.3469
V	9	4764	227	0.9528
M	2	1169	814	0.4100
Precision	0.8225	0.7909	0.7774	
F-Measures	0.4879	0.8643	0.5368	

TABLE 8. CONFUSION MATRIX OF THE WWS DATASET (ANN).

	W	V	N	Recall
W	102	57	19	0.5730
V	10	4887	103	0.9774
M	12	362	213	0.3628
Precision	0.8225	0.921	0.6358	
F-Measures	0.6754	0.9483	0.4619	

B. Patch-based Method

In the case of the patch-based method, an image augmentation node is added to increase the variability of the image data. This node applies various transformations to the image data, including setting scale factor between [0.8, 1.2], an rotation angle between [0, 3.1415] radian, aspect ratio adjustment between [1.0, 1.2], distortion between [0.0, 0.1], enabling horizontal and vertical flipping, and noise between [0.00, 0.01].

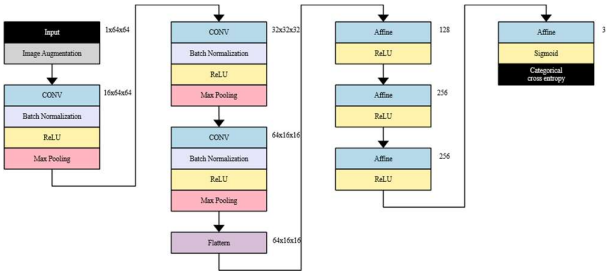


Figure 13. CNN Model for the patch-based method.

Similar to the previous case, the CNN model shown in Figure 13 can be applied to all datasets, achieving an average accuracy of 0.79. Specifically, the accuracy rates for the DOL, NIMA, VAP, and WWS datasets are 0.77, 0.82, 0.77, and 0.83, respectively.

TABLE 9. CONFUSION MATRIX OF THE DOL DATASET (CNN).

	W	V	N	Recall
W	134	20	19	0.7745
V	1	566	202	0.7360
M	3	153	650	0.8064
Precision	0.971	0.7658	0.7462	
F-Measures	0.8616	0.7506	0.7751	

TABLE 10. CONFUSION MATRIX OF THE NIMA DATASET (CNN).

	W	V	N	Recall
W	150	35	23	0.7211
V	3	670	110	0.8556
M	2	127	596	0.8220
Precision	0.9677	0.8052	0.8175	
F-Measures	0.8263	0.8296	0.8197	

TABLE 11. CONFUSION MATRIX OF THE VAP DATASET (CNN).

	W	V	N	Recall
W	89	13	13	0.7739
V	8	619	182	0.7651
M	6	169	607	0.7762
Precision	0.864	0.7727	0.7568	
F-Measures	0.8164	0.7688	0.7663	

TABLE 12. CONFUSION MATRIX OF THE WWS DATASET (CNN).

	W	V	N	Recall
W	111	8	15	0.8283
V	5	687	122	0.8439
M	3	80	372	0.8175
Precision	0.9327	0.8864	0.7308	
F-Measures	0.8774	0.8646	0.7717	

IV. Discussions

Although the average accuracy of the patch-based method is below 0.80 and lower than that of the region-based approach, it still outperforms the region-based method due to significantly improved recall rates for all non-vegetation classes. For instance, the recall rate for the non-vegetation class in the WWS dataset increased from 0.36 to 0.81, and for the water class in the VAP dataset, it rose from 0.34 to 0.77. Additionally, the patch-based model is less complex compared to the region-based method.

The differences between the results of these two methods may be due to several factors. For example, the ANN model used in the region-based method may not be deep enough to capture the complex relationships among the extracted region properties. Another factor influencing accuracy could be the number of target classes, currently set to three. While increasing the number of classes to four or more might improve results by better differentiating certain land cover types currently grouped together, it could also reduce accuracy by decreasing the number of samples or segments available for each class. In that case, a more refined feature selection process may be necessary to better separate class features and eliminate unnecessary features that could negatively influence the classification process.

V. Conclusions

This paper presents a study on the application of machine learning models for classifying water, vegetation, and non-vegetation land cover types in archived grayscale aerial imagery. The study begins by segmenting the input images using superpixel algorithms, with the extracted segments mapped to reference data manually prepared by external experts. ANN and CNN models are developed to evaluate classification performance using both region-based and patch-based methods. The region-based method achieves an overall accuracy of approximately 0.83 across all datasets, while the patch-based method achieves an average accuracy of around 0.79. Although the patch-based method shows slightly lower overall accuracy, it significantly improves recall rates, particularly for the water and non-vegetation classes.

Acknowledgment

We would like to thank the Department of Special Investigation (DSI) for preparing the aerial imagery and to the experts for their assistance in creating the ground truth maps. This study is a part of a research project funded by Thailand Science Research and Innovation (TSRI).

References

[1] T. Blaschke, “Object based image analysis for remote sensing,”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, no. 2010, 2010, pp.2–16.

[2] A.S. Laliberte, and A. Rango, “Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery,”, vol. 47, no. 3, 2009, pp.761–770.

[3] Z. Lan, and Y. Liu, “Study on Multi-Scale Window Determination for GLCM Texture Description in High-Resolution Remote Sensing Image Geo-Analysis Supported by GIS and Domain Knowledge,”, International Journal of Geo-Information, vol. 7, no. 175, 2018. pp. 1–24.

[4] B. Feizizadeh, T. Blaschke, , D. Tiede, M.H.R Moghaddam, “Evaluating fuzzy operators of an object-based image analysis for detecting landslides and their changes,”, Geomorphology, vol. 293, no. 2017, 2017, pp. 240–254.

[5] R., K. Gibson, A. Mitchell, and H.C.Chang, “Image Texture Analysis Enhances Classification of Fire Extent and Severity Using Sentinel 1 and 2 Satellite Imagery,”, remote sensing, vol. 15, no. 3512, 2023, pp. 1–21.

[6] N. Mboga, C. Persello, J.R. Bergado, and A. Stein, “Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks,”, remote sensing, vol. 9, no. 1006, 2017, pp. 1–18.

[7] M. Singh, D. Evans, B.S. Tan, and C.S. Nin, “Mapping and Characterizing Selected Canopy Tree Species at the Angkor World Heritage Site in Cambodia Using Aerial Data,”, PLoS One, vol. 10, no. 4, 2015, pp.1–24.

[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC Superpixels Compared to State-of-the-art Superpixel Methods,”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, 2012, pp. 2274–2282.

[9] M.V. Bergh, X. Boix, G. Roig, B. Capitani, and L.V. Gool, “SEEDS: Superpixels Extracted via Energy-Driven Sampling,”, In Computer Vision–ECCV 2012, Berlin:Springer, 2012, pp. 13–26.

[10] Y.J. Liu, C.C. Yu, M.J. Yu, and Y. He, “Intrinsic manifold slic: A simple and efficient method for computing content-sensitive superpixels,”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 3, pp. 653–666.