A RULE-BASED OBIA APPROACH FOR TOBACCO PLANTATION CLASSIFICATION IN SUKHOTHAI PROVINCE, THAILAND

Duy Tran Huu [1], Diep Nguyen Thi Hong [1], Minh Vo Quang [1], Andrew Makowicki [2], Pariwate Varnakovida [2]

[1] Department of Land Resource Management, Can Tho University (CTU)

[2] KMUTT Geospatial Engineering and Innovation Center (KGEO), King Mongkut's University. of Technology Thonburi (KMUTT).

E-mail: tranhuuduy0407@gmail.com

ABTRACT

The use of satellite imagery has supported the identification of crops, crop monitoring and agricultural management. Cash crop monitoring is a substantial agricultural management practice for commodity purposes in Thailand. Our study has applied a hierarchical rule-based classification schema combined with feature selection and membership functions to identify tobacco plantations in Sukhothai province, Thailand using Landsat 8 imagery. The practical classification approach was used on imagery spanning from 2014 – 2017. The overall accuracy under our rule-based classification is 85.61% with a Kappa index of 0.73 in 2017 imagery in Sukhothai. The average area of tobacco plantations from 2014 – 2017 is 4,740 hectares, representing 0.72% of the entire province. The accuracy assessment of our rule-based schema could be improved with more feature selections based on a larger sample size. However, our study successfully utilized 15m resolution satellite data in identifying current tobacco plantations to further advance the prospects of precision agriculture in Thailand. Overall, the rule-based classification approach significantly reduces the time dedicated to in field work manually identifying the area of tobacco plantations.

1. INTRODUCTION

1.1. Background

Tobacco (Nicotiana Tabacum) is an herbaceous plant. In 2011, around 4,200,000 hectares of land were devoted to tobacco growth, representing less than 1% of total arable land globally. However, in several low- and middle-income countries, the percentage of arable land devoted to tobacco growth has recently increased [2-9]. The change of tobacco planting area is the impact of climate change in each year, causing difficulties in preparing the land for cultivation. Tobacco plantations are generally located in lowland floodplains consisting of clay. The poorly drained soil characteristics of clay are not suitable for the tobacco crop. Tobacco is also effected by water scarcity and drought in the dry season [10]. Since tobacco is a sensitive crop reliant on various factors, the control of cultivated areas and insurance on sustainable development of tobacco plantations is becoming urgent.

To ensure the development of sustainable tobacco cultivation and agricultural development, practices involving land use planning and land use land cover monitoring will need to be implemented. The traditional approach in remote sensing imagery has been per-pixel classification using multidimensional spectral band analysis. In this approach, the objects in the image are grouped into categories using statistical analysis on a per-pixel basis [3], a process known as object-based image analysis (OBIA). Some of these algorithms for image classification, such as the well-known maximum likelihood classifier, assume normal distributions for data analysis, which is not always the case [1-3]. Frequently, the classification analysis ends with filtering in order to eliminate noisy pixels in the final land cover map. By introducing the concepts of neighborhood pixel, distance, and location, homogeneous pixels can be grouped into objects through a 'segmentation' of object in an image. This is the main conceptual framework behind OBIA. LANDSAT image data provides an affordable means of mapping vegetation [1] and has been widely used for crop mapping and monitoring. A study on coffee plantations in Colombia show that classification through OBIA and LANDSAT imagery produces accurate land use land cover results [1].

1.2 Vegetation Indies

The Normalized Difference Vegetation Index (NDVI) is a simple numerical indicator that can be used to analyze remote sensing measurements. NDVI is related to vegetation, where healthy vegetation reflects very well in the near infrared. Index values can range from -1.0 to 1.0 but vegetation values typically range between 0.1 and 0.7. Free standing water gives a rather low reflectance in both spectral bands and soils generally exhibit a near-infrared spectral reflectance somewhat larger than the red and also generate rather small positive NDVI values (0.1 to 0.2). A study by used MODIS imagery and statistical methods to identify and monitor tobacco plantations in Zimbabwe Over three seasons, the average tobacco yield estimates were 98% accurate [16]. Compared with the tradition method, which generally overestimated tobacco yield at 112%, the new approach performed more accurately. The Land and Water Mask (LWM) index is a tool used to differentiate between land and water. The variable has been employed to classify various waterbody types. Index values can range from 0 to 255, but water values typically range between 0 and 50 [17]. The NDVI and LWN is mostly used to classify different among land cover/land use areas.

The Visible-Band Difference Vegetation Index (VDVI) is derived based on the three bands of visible light. The values of VDVI are within $[-1, 1]$ and the accuracy of the vegetation extraction based on VDVI is higher than other visible light band-based Vegetation Indices (VI) and Green band. Furthermore, the accuracy of VDVI has been reported to be over 90% [18-19]. Brightness is an attribute of visual perception in which a source appears to be radiating or reflecting light. In the RGB color space, brightness can be thought of as the arithmetic mean (μ) of the red, green, and blue color coordinates. These two indies support NDVI in defining correctly the specific culturing from the other vegetation.

1.3 Rule-based OBIA Classification

In order to establish the distribution map of tobacco cultivation, this study establishes a rule-based classification OBIA method. This application harnesses the ability to analyze high resolution (15m) LANDSAT imagery. In the integration of GIS and remote sensing data, the use of ruled-based (also known as knowledge-based classifiers), neural networks, and expert-knowledge have been associated with higher accuracy results than statistical classifiers [4]. For instance, Murai and Omatu (1997) demonstrated that ruled-based methods, when integrated to neural networks, can improve the land cover classification by about 9% [8]. The ruled-based classification is implemented in this research through the use of a very simple decision tree. In this framework, a data set is classified according to the decision tree, and the pixel (or object) is assigned to a class according to the leaf node into which the observation falls. The decision rules in each node can be based solely on analyst expertise, which is difficult to implement across different times and geographic places, or they can be defined based on the statistical analysis of training data [20]. The classification system in this research applies a univariate decision tree in which the decision rule in each branch is defined by statistical analysis of a single feature.

1.3. Accuracy assessment

Overall accuracy (OA) is easily interpreted as it represents the percentage of classified pixels or objects that corresponds to errors of commission and omission [5-11]. While Kappa index (K) can be used to assess statistical differences between classifications. The estimate of K is based on the difference between the actual agreement (the major diagonal) and the chance agreement indicated by the row and column totals. This value is computed for each error matrix and is a measure of how well the produced classification agrees with the reference data, values range from -1 to +1 [5]. The higher the value of kappa, the better the classification performance. As values in the off-diagonal increase, the value of kappa decreases.

2. MATERIALS AND METHODS

2.1. Study Area

Sukhothai province is located in the lower northern section of Thailand, approximately 440 kilometers north of Bangkok. The province covers an area of 6,596 square kilometers (Figure 1) [6].

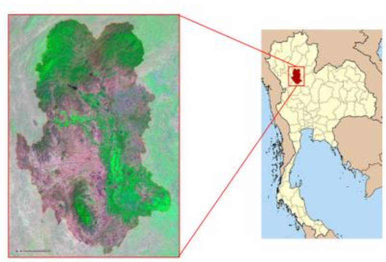

Figure 1. Sukhothai province

Tobacco plantations are mainly concentrated in Si Sisang District, Muang District, and Sawankhalok District of Sukhothai province. The total area of tobacco plantations in Sukhothai for 2009 was 8948.06 hectares, representing 1.36% of land cover within the province. In this region, tobacco-rice rotation crop covers 83% of the tobacco growing areas [6-10]. Approximately 10,000 people (6,000 tobacco companies) produce tobacco leaves for tobacco companies and export companies [10]. There is a high demand for tobacco in Sukhothai Province, Thailand. Muang district has the most fertile soil which is sutable for tobacco cultivating with 3411 hectares [10].

2.2. Materials

Methods of object-based classification were studied using multispectral LANDSAT 8 from 2014 – 2017. The imagery covers the Sukhothai province and was obtained at a spatial resolution of 30m (Table 1) from the United States Geological Survey (USGS) (**Error! Reference source not found.**). Satellite imagery was collected for the month of February, when tobacco plants are in full bloom and easily identifiable.

Table 1. Image data used in this study

No	Landsat Scene ID	Date	Clouds Ratio	Path/Row	File format
1	LC81300482014040LGN01	09/02/2014	7.42%	130/048	GeoTIFF
2	LC81300482015059LGN01	28/02/2015	0.04%	130/048	GeoTIFF
3	LC81300482016046LGN01	15/02/2016	0.09%	130/048	GeoTIFF
4	LC81300482017048LGN00	17/02/2017	0.31%	130/048	GeoTIFF

(Source: http://earthexplorer.usgs.gov)

In addition to the 2016 land use map of Sukhothai province, Google Earth's Street View (Pegman) function was used to collect data for land cover type in Sukhothai. A total of 815 survey points were collected: 124 points for Agricultural, 150 Forest, 125 Urban and Bare Soil, 124 Water, 180 Rice, and 112 Tobacco points. The surveyed points of Google Earth were used for building the classification "key" and used to select known sample points in rule-based classification.

2.3.Methods

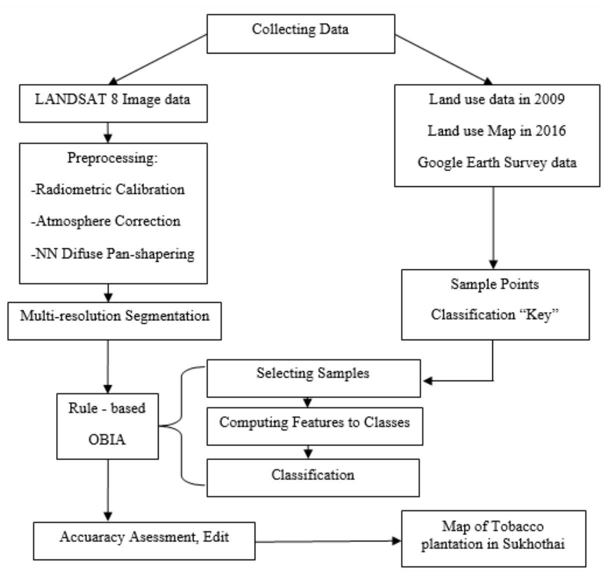

Figure 2. Flowchart of land cover classification of tobacco plantations in Sukhothai

2.4 Rule–based Classification

Step 1: Define classes and Class features

Four main classes were defined for this study: Agriculture, Forest, Urban – Bare Soil, and Water. Furthermore, Tobacco and Rice were established as sub-classes for Agriculture. Six classes and their subsequent inheritance relationships were established in the class hierarchical window of eCognition. The GeoTIFF file was modified to support a smaller study area and reduce the processing time. All of the necessary classes defined in the sample selection were present in the study area. After creating the inheritance relationship among 6 classes shown in Figure 3, class features were defined (Table 2). These features were allocated to the Standard Nearest Neighbor Feature Space to build the chart parameters for each object, based on the sample selection in Step 3.

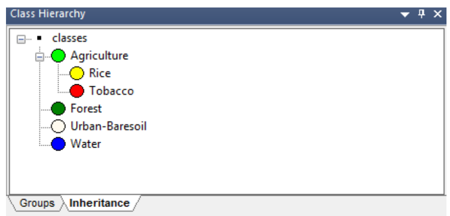

Figure 3. The inheritance relationship of class in this study

Table 2. The class features used in this study

Features name	Formula	LANDSAT 8
NDVI	NDVI = (NIR-Red)/(NIR+Red)	(Mean Band 5 - Mean Band 4) /(Mean Band 5 + Mean Band 4)
VDVI	VDVI = (2∗ Green− Red −Blue)/(2*Green +Red +Blue)	(2* Mean Band 4- Mean Band 3- Mean Band 2)/(2* Mean Band 4+ Mean Band 3+ Mean Band 2)
LWN	LWN= (MIR) / (Green)	Mean Band 6 /Mean Band 3
Brightness	μ =(Red+Green+Blue)/3	(Mean Band 4+ Mean Band 3+ Mean Band 2)/3

Step 2: Multi-resolution Segmentation

Image segmentation was carried out on the pan-sharpened imagery. Multiresolution segmentation (MRS), as implemented in Trimble eCognition software, was used for generating the image objects. MRS is a region-merging pair-wise algorithm in which each data point is associated with another data point, where the delineation algorithm will predict which of the two is "more" relevant according to the input parameters. It segments on pixel or object level based on user-defined scale, shape, and compactness parameters [5]. The multi-resolution function was applied for the extraction of image objects using modifiable scale parameters, single layer weights, and homogeneity concerning color and shape [20]. Within the shape setting, smoothness and compactness can also be weighted from 0 to 1. In our study, the new process is created in the Process Tree, we select the Multi-resolution Segmentation Algorithm for processing in Level 1, segmentation parameter used for analysis applying to do the Multi-resolution Segmentation process is shown in Table 3.

Table 3. Segmentation parameter used for analysis

Segmentation Level	Scale	Active Bands	Smoothness	Compactness
Level 1	15	1 - 7	0.5	0.5

Step 3: Sample Selection

The survey points on Google Earth help formed the key image interpretation (**Error! Reference source not found.**). Sample features were manually selected using the Select Sample tool. We selected 50 - 100 samples for each class.

Step 4: Rule-Based Classification

We applied the Membership Function to obtain the suitable feature values from Sample Editor for each class. The values were obtained in the Class Description window (Figure 4) and compared with the other classes. We selected the features which have less overlap from one another. Range of the feature's value were selected and also changed based on the features statistical value displayed in Sample Editor window. The suitable values to specify each class based on the statistical analysis were selected for the development of Membership functions. For the main classes, we use the Classification process in Process tree with the Active Class which needed to classify. For the classification of the sub-classes Tobacco and Rice, a class filter was added to Agriculture.

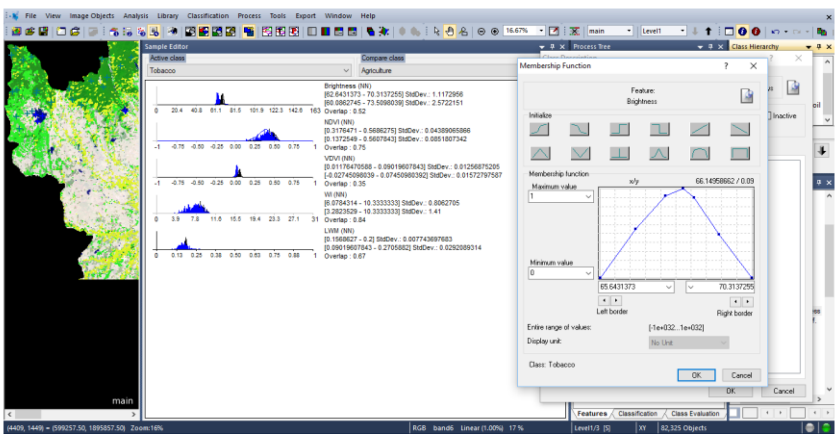

Figure 4. Applying Membership function to rule-based classification

All these processing methods have been applied to the LANDSAT 8 imagery from 2014 to 2017 for classifying tobacco plantations in Sukhothai province.

2.5. Accuracy Assessment

The accuracy assessment was based on field data collected for land use land cover of Sukhothai in January 2017. The data comprised of 82 known tobacco plantation points and 50 non-tobacco plantation points to calculate the accuracy of the rule-based classification.

3. RESULTS

3.1. Image preprocessing

Figure 5 shows the satellite imagery of the site selection before and after the preprocessing steps. The right image has gone through the preprocessing steps of radiometric calibration, eliminating the atmosphere factors (Dark object subtraction) and improved spatial resolution. After-preprocessing, the imagery is exported as a TIFF file with the projection UTM 47 North. Comparing with the original raw data, the data after preprocessing with shades brighter and clearer makes the ground objects easily distinguishable.

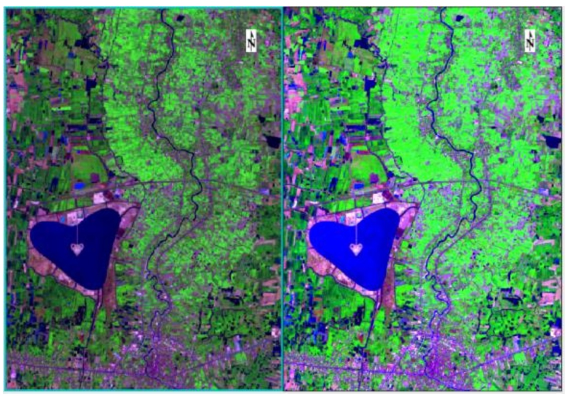

Figure 5. Comparing the image data before (left) and after (right) preprocessing

3.2. Rule-based OBIA Classification

Land cover classification has process by the rule-based methods applied the range of feature's value shown in Table 4. Four main classes and two sub-classes are classified by using the normal classification process in the Process Tree. After all classification process, the result is export as polygon feature shapefile ready for accuracy assessment and mapping.

Table 4. The range of features value used for classification

Features	Agriculture	Forest	Urban-Bare Soil	Water	Rice	Tobacco
NDVI	0.137 - 0.581	0.098 -0.553	-0.121 – 0.176	-0.607 – 0.145	0.161-0.524	0.317 - 0.569
VDVI	-	-0.275 – 0.0274	-	-	-0.004-0.067	0.012 - 0.09
LWN	-	-	-	0.0078-0.157	-	-
Brightness	-	55.611-70.953	-	62.643-70.313	-	65.643-70.314

3.3. Accuracy assessment

The confusion matrix shown in 5 provides the overall accuracy of the classification results. Overall accuracy was 85.61% and the Kappa statistic was 73.24%. Accuaracy indices are acceptable in classifying tobacco plantations.

3.4. Tobacco plantation in Sukhothai province

In Figure 6, Urban-Bare soil was classified as having the largest area, covering 40.05% of Sukhothai province. Tobacco covers 0.72% of Sukhothai. Rice plantations cover 8.81% of Sukhothai province. The concentration of rice crop is mainly within the central region of Sukhothai. Water covers 5.04% of the province. This includes the Yom Rivers, lakes and irrigation networks throughout the province (Figure 7).

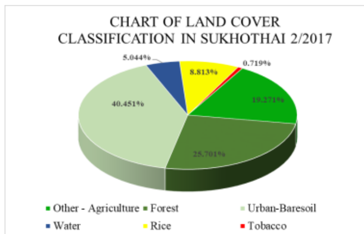

Figure 6. Chart of Sukhothai land cover classification 2/2017

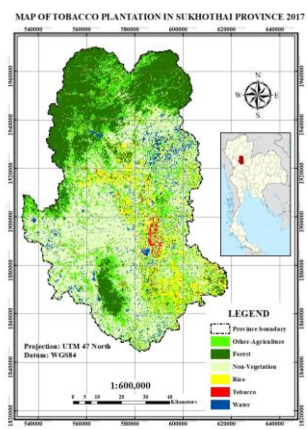

Figure 7. Map of tobacco plantation in Sukhothai province 2017

To calculate the tobacco crops area in Sukhothai, we use the feature's value shown in Table 5.

Table 5. The Tobacco class's features used for LANDSAT 8 image analysis

Class Feature	2014	2015	2016	2017
NDVI	0.347-0.574	0.117-0.677	0.119-0.778	0.317 - 0.569
VDVI	0.009-0.096	0.004-0.105	0.006-0.115	0.012 - 0.09
Brightness	72.804-93.107	69.981-88.102	68.524-80.001	65.643-70.314

Figure 8 shows the distribution of tobacco plantations classified annually in February from 2014 to 2017. The average area of tobacco plantation in Sukhothai province is classified as 5270 ha, accounting for 0.80% of the entire area of Sukhothai Province. In which area is the largest classification in 2015 to 6115,275 hectares, accounting for 0.927% of the area of Sukhothai province. In Figure 9, classification results over the years shows that tobacco plants concentrated mainly in Si Sam Rong and Muang districts with high density.

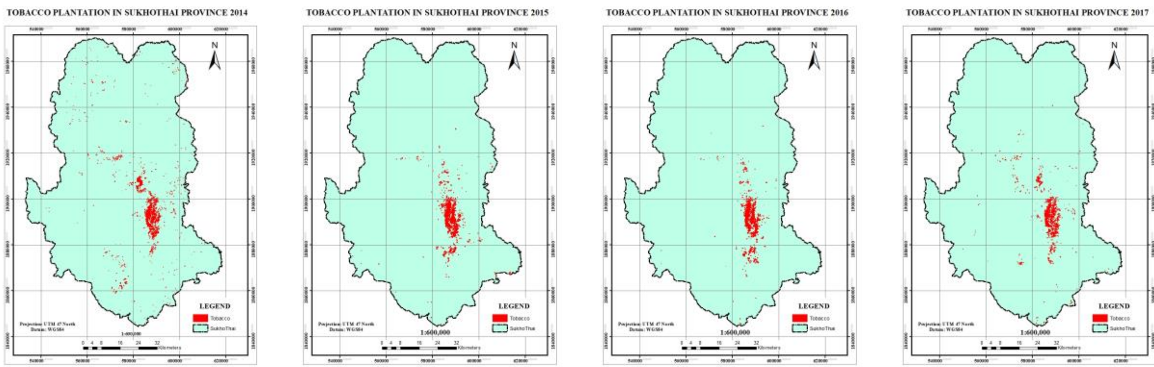

Figure 8. Tobacco plantation area classified in Sukhothai 2014 - 2017

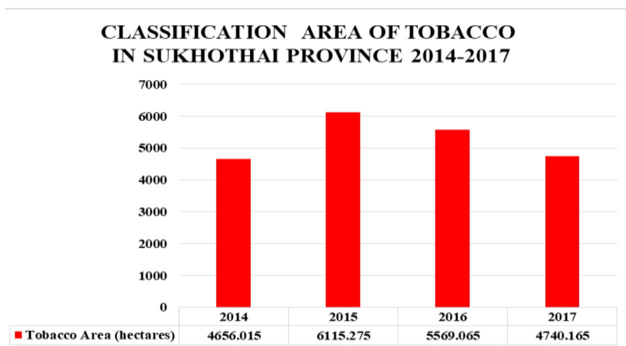

Figure 9. Chart of classification areas of tobacco in Sukhothai province 2014 - 2017

4. CONCLUTION

A rule-based classification approach using remote sensing techniques and OBIA were used to classify and map tobacco plantations in Sukhothai province. The rule-based approach can be applied to serve as a tool in monitoring agricultural landscapes, crop management such as forecasting pest influences, and projecting the impact of climate change on various crop types. In this study, we have successfully applied data from LANDSAT and methods incorporating a rule-based OBIA classification approach towards identifying tobacco plantations and determining the area of tobacco agricultural landscapes from 2014 to 2017 within Sukhothai province. The rule-based classification scheme built for this project can be used to monitor tobacco crop through the automation of land use land cover identification using remote sensing and OBIA techniques. The schema improves upon agricultural management while also supporting environmental-sustainability.

5. REFERENCES

[1] Bolanos, S. (2007). Using Image Analysis and GIS for coffee Mapping, 119.

[2] C.G Warford Lock, F. L. S. (Ed.). (1886). Tobacco: Growing, curing, and manufacturing. University of California.

[3] De Jong, S. and Van der Meer, F. (2004). Remote Sensing Image Analysis: Including the Spatial Domain. Kluwer Academic Publishers.

[4] Fierens, F., Kanellopoulos, I., Wilknson, G., Megier, J. (1994) Comparison and visualization of feature space behaviour of statistical and neural classifiers of satellite imagery. Geoscience and Remote Sensing symposium IGARSS '94.

[5] Kane, J., Kemp, J., & Niekerk, A. Van. (2016). Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques.

[6] Land Development Department, (2009) & (2016). Sukhothai province land use data 2009 and Sukhothai's land use map 2016. Website: http://www.ldd.go.th/

[7] Lewinski, S., Bochenek, Z., St., L., & Bochenek, Z. (2008). Rule-based classification of SPOT imagery using object-oriented approach for detailed land cover mapping. *Proceedings of the 28th EARSeL Symposium „Remote Sensing for a Changing Europe" 2-5 June 2008, Istanbul, Turkey*, (June), 1–8.

[8] Murai, H., Omatu, S. (1997) Remote Sensing Image Analysis using a neural network and knowledge-based processing. International Journal of Remote Sensing, 18(4), 811- 828

[9] Novotny, T. E., Bialous, S. A., Burt, L., Curtis, C., da Costa, V. L., Iqtidar, S. U., … Tursan d'Espaignet, E. (2015). The environmental and health impacts of tobacco agriculture, cigarette manufacture and consumption. *Bulletin of the World Health Organization*, *93*(12), 877–80.

[10] Office of Natural Resourse and Enviroment Policy and Planning Thailand, (2014). "A summary of action plan for the problem of tobacco farmers in Sukhothai Province".

[11] Radoux, J., & Bogaert, P. (2017). Good practices for object-based accuracy assessment. Remote Sensing, 9(7). https://doi.org/10.3390/rs9070646.

[12] Rahman, M. Z. A. (n.d.). Characteristics of remote sensing data & pre-processing, (Sghg 1473), 32.

[13] Sun, W., Chen, B., & Messinger, D. W. (2014). Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images. *Optical Engineering*, *53*(1), 13107. https://doi.org/10.1117/1.OE.53.1.013107.

[14] Sun, W., Chen, B., & Messinger, D. W. (2014). Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images. *Optical Engineering*, *53*(1), 13107.

[15] Survey, U. S. G. (2005). Landsat 8 (L8) Data Users Handbook. *America*, *8*(1993), 1993–1993. https://doi.org/http://www.webcitation.org/6mu9r7riR

[16] Svotwa, E., Masuka, A. J., Maasdorp, B., & Murwira, A. (2014). Estimating Tobacco Crop Area and Yield in Zimbabwe Using Operational Remote Sensing and Statistical Techniques, *2*(5), 84–91.

[17] Uddin, K. (n.d.). Hands on Exercise Using eCognition Developer, 1–58.

[18] X. Wang,M. Wang, S. Wang, and Y.Wu, "Extraction of vegetation information fromvisible unmanned aerial vehicle images," Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, vol. 31, no. 5, pp. 152–159, 2015.

[19] Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*.

[20] Yuan, F., & Bauer, M. E. (2006). Mapping impervious surface area using high resolution imagery: A comparison of object-based and per pixel classification. *American Society for Photogrammetry and Remote Sensing Annual Conference Proceedings 2006, Reno, Nevada*, (2003).