The Application of Unmanned Aerial Vehicles (UAVs) and Extreme Gradient Boosting (XGBoost) to Crop Yield Estimation: A Case Study of Don Tum District, Nakhon Pathom, Thailand

Kulpanich, N.,[1*] Worachairungreung, M.,[1] Thanakunwutthirot, K.[2] and Chaiboonrueang, P.[3]
[1]Suan Sunandha Rajabhat University, Faculty of Humanities and Social Sciences, Geography and Geo-Informatics Program, Bangkok, Thailand, 10300, E-mail: nayot.ku@ssru.ac.th,* morakot.wo@ssru.ac.th
[2]Suan Sunandha Rajabhat University, Faculty of Fine and Applied Arts, Digital Design and Innovation Program, Bangkok, Thailand, 10300, E-mail: kunyaphat.th@ssru.ac.th
[3]Thaicom Public Company Limited, Remote Sensing Data Analytics Office, 349, 28th Floor, SJ Infinite 1 Business Complex, Vibhavadi Rangsit Road, Chompol, Chatuchak, Bangkok 10900, Thailand
E-mail: prin_cha@thaicom.net
Corresponding Author

DOI: https://doi.org/10.52939/ijg.v19i2.2569

Abstract
Rice (Oryza sativa L.) is a staple food for more than half of the global population. This research, therefore, aims to explore the estimation of crop yields towards the application of unmanned aerial vehicles (UAVs). The research areas are the sample rice fields owned by Sam Ngam Large-Scale Rice Production Community Enterprise in Don Tum District, Nakhon Pathom. The data collected by both RGB and multispectral UAVs was used for estimating the crop yields of Rice Department 41 (RD41), a rice variety, and then analyzed by a geographic information system (GIS). Multiple Linear Regression was applied to factor analysis for the purpose of crop yield estimation based on the factors investigated and obtained by the UAVs. These factors included vegetation indexes (i.e. Normalized Difference Vegetation Index, Green Normalized Difference Vegetation Index, and Triangular Greenness Index), plant height, and canopy coverage. The prediction of the analysis model was proved to be valid (R^2 = 0.99; RMSE = 2.506 g.). Extreme Gradient Boosting (XGBoost) was applied to increase the accuracy of the estimation (RMSE = 0.557 g.; MAE = 0.364). The findings of study showed that the utilization of UAVs could contribute to the estimation of crop yield in the research areas.

Keywords: Crop Yield Estimation, Extreme Gradient Boosting, Multispectral Camera, Rice Unmanned Aerial Vehicles

1. Introduction

Rice (Oryza sativa L.) is an essential source of energy for human beings. It is a staple food for more than 50% of the world's population by which approximately 20,000 tons of rice are consumed each year [1] and [2]. However, the demand for rice has been tremendously rising [3]and [4]owing to myriad global concerns such as overpopulation [5], food insecurity [6], and climate change [7] Countries in Southeast Asia, a major rice-producing region, such as Vietnam, Myanmar, and Thailand are widely acknowledged as primary rice exporters across the globe. In particular, the Chao Phraya River Basin [8], the largest basin of Thailand, is one of the most intensive rice-producing areas worldwide. In 2019, around 40% of the total labor force in Thailand belonged to the agricultural sector, and 6.04 million tons of rice (with a value of 3.74 billion USD) were exported, making the country become the world's second largest rice exporter of that year [9]. Therefore, it can be assumed that rice is a crucial economic plant in Thailand which is not only consumed domestically but also globally. Unfortunately, even though Thailand's economic performance is highly correlated with rice production outputs, the Office of Agricultural Economics [10] reports that the national rice yields have declined and become uncertain over the past five years. As a result, an accurate estimation of rice production is necessary in order to maximize the quantity and quality of each paddy field.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

At present, the advancement of geospatial technology is flourishing, especially the application of remote sensing and satellite imagery in agriculture [11] [12] and [13] According to González et al., [14] 20m spatial resolution imagery provided by the Sentinel-2 satellites was employed to monitor rice fields and their outputs and record data in time series based on 4 indexes: NDVI, NDWIMF, NDWIGAO, and BSI. The study shows that NDWIMF could estimate crop yields most effectively. Another study conducted by Hashemi et al., [15] reveals that the Synthetic Aperture Radar (SAR) data from Sentinel 1 could predict the crop yields of rainfed paddy fields efficiently. Several scholarly works have explored the application of unmanned aerial vehicles (UAVs) to crop yield estimation. For instance, Reza et al., [16] utilized low-altitude UAVs to predict crop yields by means of the k-means clustering technique. Ge et al., [17] investigated rice production by analyzing plant nitrogen concentration (PNC) which signified the correlation between crop yields and nitrogen content in agricultural fields based on the photos taken by UAVs. Phan and Takahashi [18] applied a UAV LIDAR system to examine height of rice plants. They discovered a positive relationship between the height of rice plants and the crop yields observed by LIDAR's mineral exploration. Their study shows that the UAV LIDAR system is useful for tracking the rice growth, and geoinformatics can benefit contemporary agriculture. However, the limitation of the satellite image is the lower spatial resolution than UAV images. In addition, using hyperspectral satellite images is possible, but at the same cost of operation as LIDAR UAVs. Notwithstanding, UAVs are cheaper due to their variety and better spatial resolution than satellite imagery. UAVs are cheaper because they are more versatile and have better spatial resolution than satellite images, making them more accessible to the public, and geospatial methods can be applied to estimate yield.

Therefore, this research attempts to scrutinize factors affecting rice production towards the use of unmanned aerial vehicles (UAVs) and then estimate crop yields of the sample rice fields in Don Tum District, Nakhon Pathom based on the explored factors. The expected outcomes of this study are increases in rice productivity and rice farmers' revenues by virtue of the UAV technology.

2. Materials and Methods

2.1 Study Area

Nakhon Pathom is a province in the central region of Thailand where latitude and longitude coordinates are 13.814029 and 100.037292. It is a province in the Thachin River Basin, and its size is 2,168.327 km². Distance between Bangkok and Nakhon Pathom is 60 kilometers. Most areas of the province are plain fields with neither mountains nor forests, which are favorable to rice production. The central areas are low plains composed of plateaus and water sources. The eastern and southern areas are low plains along the Thachin River where many natural and artificial canals can be found.

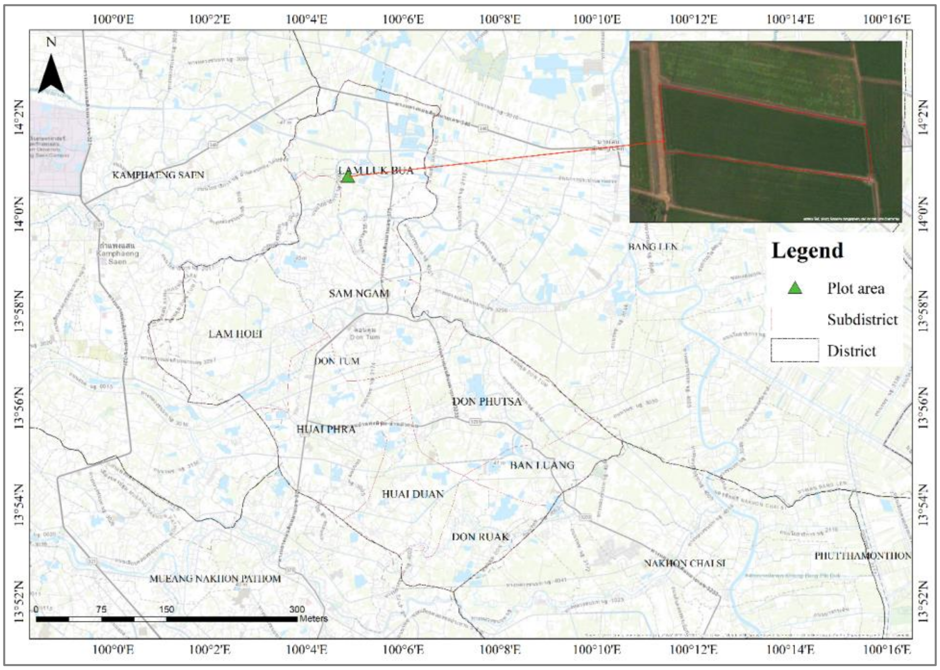

Figure 1: Study area

However, the northern and northeastern areas of the province are highlands which are not suitable for rice farming [19]. According to the Land Development Department [20], the majority of land in Nakhon Pathom is utilized for agriculture. The land area of 680.18 km² belongs to rice fields, accounting for 46.17% of the total agricultural land (1,472.99 km²). This fact shows that rice is a significant economic plant of Nakhon Pathom. Thus, in this study, the rice fields of Sam Ngam Large-Scale Rice Production Community Enterprise in Don Tum District, Nakhon Pathom, were selected as sample fields. Sam Ngam Large-Scale Rice Production Community Enterprise, consisting of 40 member households, owns an agricultural land of 1 km². It is located in Koh Tan Village, Moo 11, Sam Ngam Sub-district, Don Tum District, Nakhon Pathom. The village occupies an agricultural land of 3.2 km² (see Figure 1), and most of the working population in the village are farmers.

2.2 Data Collection

This research applied field experimentation in agriculture by which data was obtained from two connected sampled rice fields with the size of 0.016 km² each. The output data from the sample fields was collected from 20 plots scattering across the fields (1m high x 1m wide each). The sub-fields or spots were selected by UAVs when the farmers trapped water in the fields. Vegetation indexes were employed to classify the field areas into three groups: soil, water, and vegetation. Next, the samples were selected from all of the three groups. In this study, aerial exploration and data collection towards the use of UAVs were allowed by the rice field owners. Accordingly, a DJI Phantom 4 Pro V2.0 camera drone was utilized because of the following features: satellite positioning systems (GPS/GLONASS), a 7-km flight range, a 500-m max service ceiling, wind speed resistance of 36 km/h, operating frequency between 2.40–5.85 GHz, a 4K/20 MP 1-inch (12.8mm x 9.6mm) CMOS sensor camera, FOV 84° lens with an aperture of f/2.8 - f/11 and auto focus at 1 m, and a 3-axis (pitch, roll and pan) stabilized gimbal [21] and [22]. A DJI Phantom 4 Multispectral mapping drone was also used as it was equipped with six cameras, including 1 RGB camera and other 5 cameras filtered by five different colors (red edge, near-infrared, green, red, and blue) [23].

The DJI Phantom 4 Pro V2.0's photogrammetry was operated in the double grid mode at 20 meters above the ground level. The image overlap was set up (70% side lap and 80% overlap) in order to obtain the ground sample distance of 0.58 cm/pixel. Meanwhile, the DJI Phantom 4 Multispectral's photogrammetry was also operated in the double grid mode at 20 meters above the ground level. However, its image side lap and overlap were set at 65% and 75%, respectively, in order to acquire the ground sample distance of 1.1 cm/pixel. Four ground control points were located at each corner of the connected fields. Other four ground control points were installed at each edge of the fields. In total, there were 8 ground control points for each drone's operation (as Figure 2 and Table 1).

2.3 Methods

In terms of data analysis, the concept of UAV photogrammetry was applied to UAV imagery. Orthorectification was processed in order to solve problems of relief displacement and tilt displacement by using high and low spatial data based on the Universal Transverse Mercator (UTM) which provided adjusted images of objects and landscapes as well as their locations, sizes, and shapes. The adjustment based on such concept increased the accuracy of mapping. Grounded on the literature review, this study investigated 7 factors affecting UAVs' crop yield estimation, which included Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation (GNDVI), Triangular Greenness Index (TGI), crop canopy, soil surface, plant height, and canopy cover (as Table 2).

Table 1: UAV specifications

Specifications	**DJI Phantom 4 Pro V2.0**	**DJI Phantom 4 Multispectral**
Satellite positioning systems	GPS/GLONASS	GPS/GLONASS
Flight range	7 km.	7 km.
Max service ceiling	500 m.	500 m.
Wind speed resistance	36 km/h	36 km/h
Operating frequency	2.40–5.85 GHz	2.40–5.85 GHz
Sensor and Camera	4K/20 MP 1-inch (12.8mm x 9.6mm) CMOS	Six cameras, including 1 RGB camera and other 5 cameras colors (red edge, near-infrared, green, red, and blue)
Aperture	f/2.8 - f/11	f/2.8 - f/11
Gimbal	3-axis (pitch, roll and pan) and stabilized gimbal	3-axis (pitch, roll and pan) and stabilized gimbal

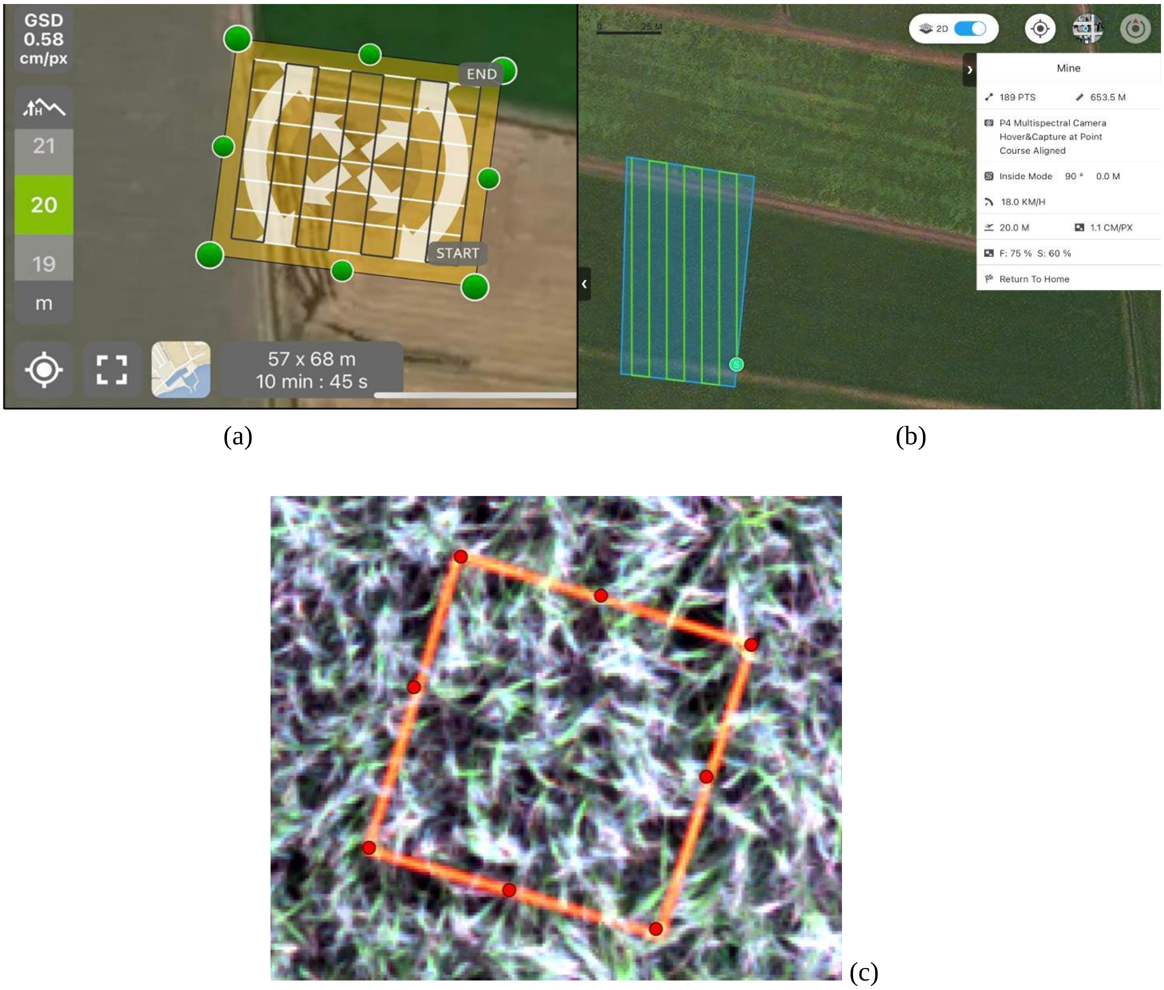

Figure 2: Data collection towards the use of RGB and multispectral UAVs. (a) Flight path of RGB UAVs (b) Flight path of Multispectral UAV (c) Ground control points in each plot area

Table 2: Factors affecting crop yield estimation

Factors	Equation	Sources
NDVI	$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$	Kong et al., [24]
GNDVI	$GNDVI = \frac{(NIR - Green)}{(NIR + Green)}$	Basso et al., [25]
TGI	$Green - (0.39\ x\ Red) - (0.61\ x\ Blue)$	Hunt Jr. et al., [26]
Crop Canopy	$Digital\ Surface\ Model$	Wan et al., [27]
Soil Surface	$Digital\ Terrain\ Model$	Wan et al., [27]
Plant Height	$Crop\ Canopy\ -\ Soil\ Surface$	Malambo et al., [28]
Canopy Cover	$CC = \frac{no.\ of\ crop\ pixels\ in\ AOI}{overall\ number\ of\ pixels\ in\ the\ AOI}$	Hunt et al., [29]

However, the plant height factor was calculated using the raster calculator in the GIS software.Next, the relationships between relevant factors were explored in order to predict the crop yields of the sample rice fields. There were 7 factors used as independent variables, while the crop yields were the dependent variable. The crop yield prediction was analyzed by Multiple Linear Regression (MLR) [30], as explained by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_i x_i + \varepsilon$$

Equation 1

Where as:

β = Estimated regression coefficient

x = Factors affecting crop yields of sample rice fields

y = Outputs of sample rice fields

In addition, this study used the Extreme Gradient Boosting (XGBoost) model to predict yield using the same factors as MLR to compare yield predictions across the sample area. Extreme Gradient Boosting (XGBoost) was learning from parameters in the same way as the regression tree family approach. This model was based on the gradient descent direction of the loss function of the last established model [31], as explained by the following equation:

$$L(\emptyset) = \sum_i l\,(\hat{y}_i, y_i) + \sum_k \Omega(f_k)\ \ \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

Equation 2

Where as:

$l(\hat{y}_i, y_i)$ = Training loss,

$\Omega(f_k)$ = The complexity of trees

f_k = The regression trees

T = The number of leaves in the tree

w = The score of the regression tree node

Then, the results from the MLR and XGBoost analysis were compared to the crop yields of each sample plot observed by field exploration and validated by the coefficient of determination (R^2). The R-Squared represents the relationship between independent and dependent variables, while the Root-Mean-Square Error (RMSE) and the Mean of Absolute value of Errors (MAE) represent the relationship between predicted crop yields and observed crop yields of 20 sample plots [32] and [33], as illustrated by the following equation:

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{pre})^2}{\sum(Y_{obs} - Y_{ave})^2}$$

Equation 3

$$RMSE = \sqrt{\frac{1}{N}\sum(\boldsymbol{Y_{obs}} - \boldsymbol{Y_{pre}})^2}$$

Equation 4

$$MAE = \frac{1}{N}\sum\left|\boldsymbol{Y_{obs}} - \boldsymbol{Y_{pre}}\right|^2$$

Equation 5

Where as:

$\boldsymbol{Y_{obs}}$ = Crop yield observed by field exploration

$\boldsymbol{Y_{pre}}$ = Predicted crop yield

$\boldsymbol{Y_{ave}}$ = Average of crop yields observed by field exploration

$\boldsymbol{N}$ = Number of data points

Finally, the crop yield estimation results were tested by Extreme Gradient Boosting (XGBoost), a tree-based machine learning algorithm. In this algorithm, decision trees are sequentially formulated and continuously learn from the errors of previous decision trees until a stopping criterion is met, gradually increasing prediction accuracy [34].

3. Results

3.1 Analysis of Factors Affecting Sample Fields' Crop Yields by UAVs

According to the exploration of the 20 plots (1 m^2 each), the multispectral UAV could analyze 3 factors, including Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation (GNDVI), and Triangular Greenness Index (TGI). The NDVI and GNDVI reflected the vegetation index values ranging from -1.0 to 1.0. Meanwhile, the TGI demonstrated the chlorophyll content of plants ranging between 15–85 mg/cm^2 [29]. It can be assumed that the higher index values, the greater amount of chlorophyll (as Figure 4). Moreover, there were two other factors analyzed by the RGB UAV, which included plant height and canopy cover. The RGB UAV performed in a double grid pattern, achieving UAV-captured RGB images which could be used for calculating the height data of the crop canopy and soil surface based on a geographic information system (GIS). The data gained from the previous process was analyzed to calculate the plant height in the sample fields. Regarding the canopy cover, the images were interpreted and divided into two types of pixel images representing rice areas and non-rice areas, respectively. The data was then calculated by a canopy cover estimation formula (as Figure 5). As a result, there were 5 factors affecting crop yields of the sample fields in total, as listed in Figure 6.

3.2 Estimation of Sample Fields' Crop Yields Based on Data Collected by UAVs

To estimate the crop yields of the sample fields, crop yields were observed from the 20 plots in order to compare them with the crop yields predicted by Multiple Linear Regression based on the following factors: Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation (GNDVI), Triangular Greenness Index (TGI), plant height, and canopy cover.

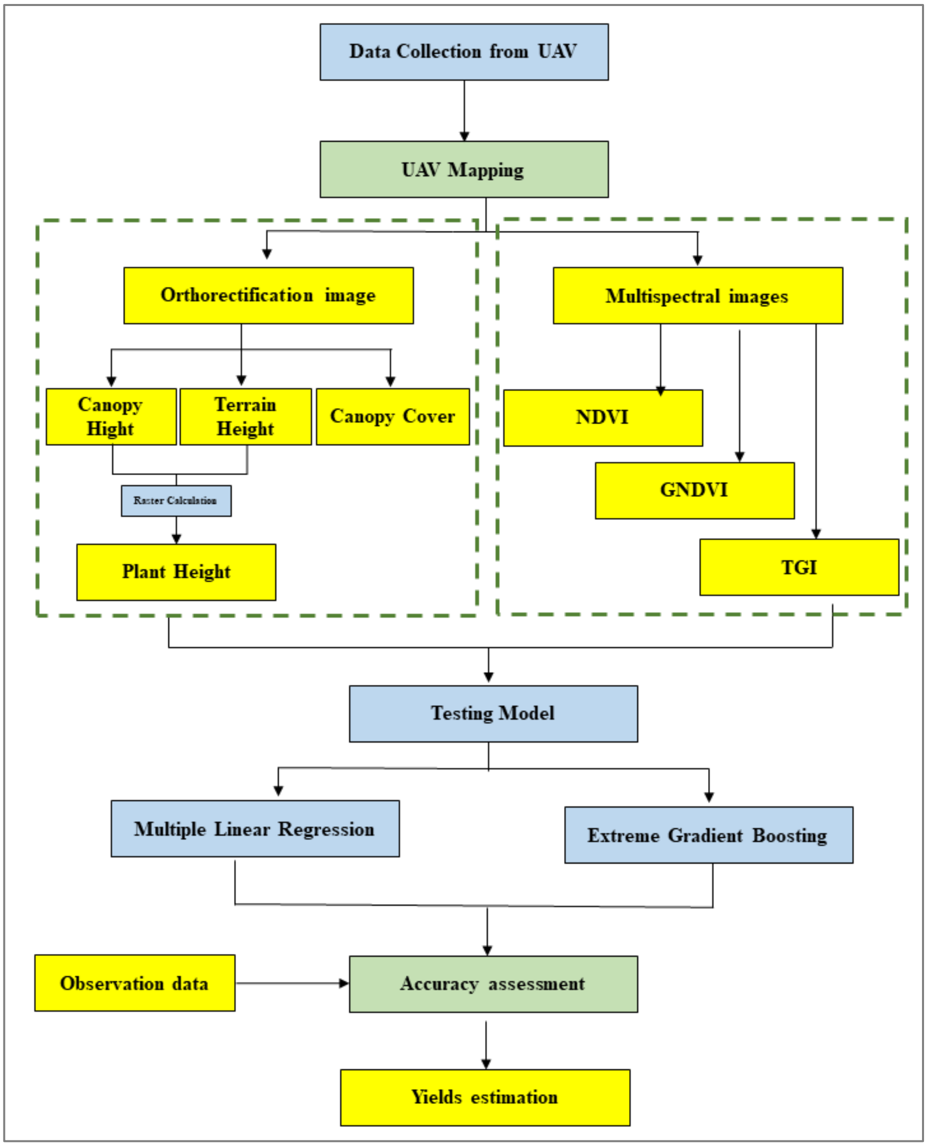

Figure 3: Conceptual framework

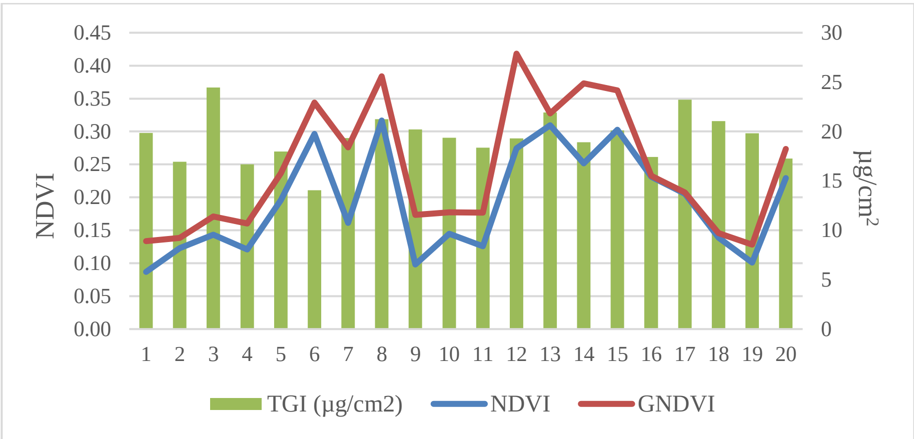

Figure 4: Analysis of data received from multispectral UAV

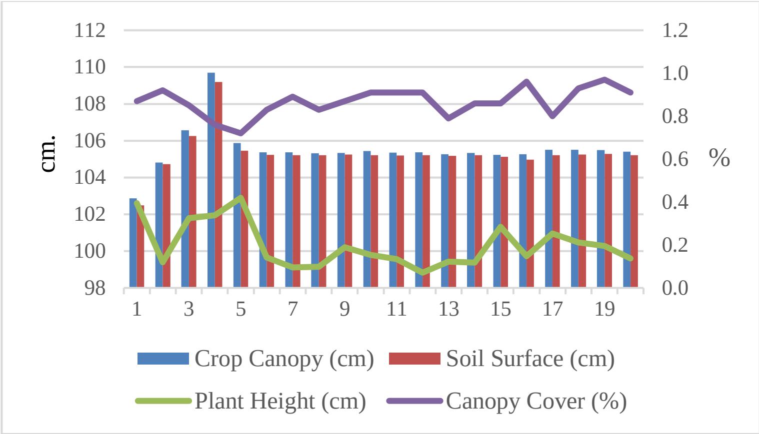

Figure 5: Analysis of data received from RGB camera from UAV

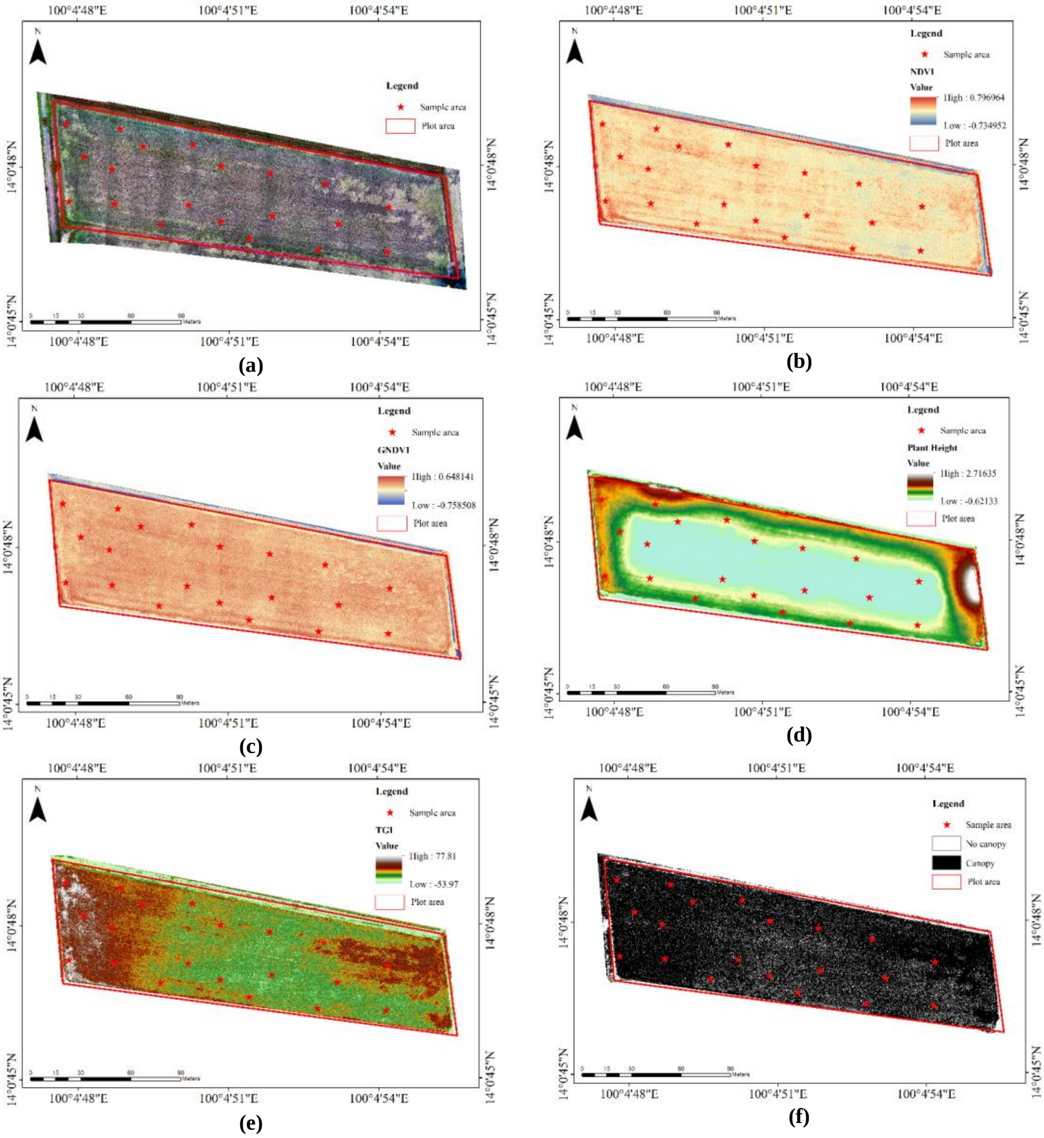

Figure 6: Factors affecting crop yields of sample fields. (a) Plot areas (b) NDVI (c) GNDVI (d) Plant height (e) TGI (f) Canopy cover

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

According to the research results, the most significant factors included canopy cover, plant height, and Triangular Greenness Index (TGI), followed by Green Normalized Difference Vegetation (GNDVI) and Normalized Difference Vegetation Index (NDVI). This study found that all factors are important for rice yield estimation and could be ordered as follows: Canopy cover, NDVI, Plant height followed by GNDVI, and TGI (as the value column in Table 3). Based on the aforementioned factors, the crop yields of the sample fields were estimated by validating the crop yields observed from the 20 plots of the sample fields by the coefficient of determination (R^2), Root-Mean-Square Error (RMSE), and Mean of Absolute value of Errors (MAE) as illustrated in Figure 7 and Table 4 respectively. In the Figure 7, the research results showed that the R-Squared value was 0.99, indicating that the prediction of the Multiple Linear Regression (MLR) model was valid. Then, the average difference between the predicted crop yields and the observed crop yields was measured by the Root-Mean-Square Error (RMSE = 2.506 g.), and Mean of Absolute value of Errors (MAE) = 1.778.

Table 3: Level of significance of factors affecting sample field's crop yields

Source	**Value**	**Standard Error**	**t**	**Pr > |t|**	**Lower Bound (95%)**	**Upper Bound (95%)**	***p*-values Signification Codes**
Intercept	0.000						
NDVI	684.102	18.571	36.836	**<0.0001**	644.518	723.685	***
GNDVI	89.649	15.500	5.784	**<0.0001**	56.611	122.687	***
TGI (μg/cm^2)	10.436	0.233	44.706	**<0.0001**	9.938	10.933	***
Plant height (m)	443.095	6.133	72.242	**<0.0001**	430.022	456.168	***
Canopy cover (%)	1,102.962	4.653	237.067	**<0.0001**	1093.046	1112.879	***

*Signification Codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 <. < 0.1 < ° < 1*

Table 4: Prediction of crop yields of 20 plots in sample fields by multiple linear regression model

Plot	**Observed Yield (g.)**	**Predicted Yield (g.)**	**Residual (g.)**
1	1,412.480	1,413.210	-0.730
2	1,343.643	1,340.669	2.974
3	1,449.320	1,450.202	-0.882
4	1,259.920	1,259.276	0.644
5	1,324.450	1,322.813	1.637
6	1,360.260	1,358.392	1.868
7	1,362.200	1,360.690	1.510
8	1,434.660	1,432.012	2.648
9	1,336.230	1,336.873	-0.643
10	1,386.880	1,388.677	-1.797
11	1,355.260	1,357.133	-1.873
12	1,459.950	1,461.910	-1.960
13	1,397.700	1,395.652	2.048
14	1,402.782	1,403.739	-0.957
15	1,522.840	1,523.634	-0.794
16	1,480.020	1,484.996	-4.976
17	1,393.400	1,395.848	-2.448
18	1,447.740	1,447.523	0.217
19	1,448.440	1,443.917	4.523
20	1,425.520	1,425.953	-0.433

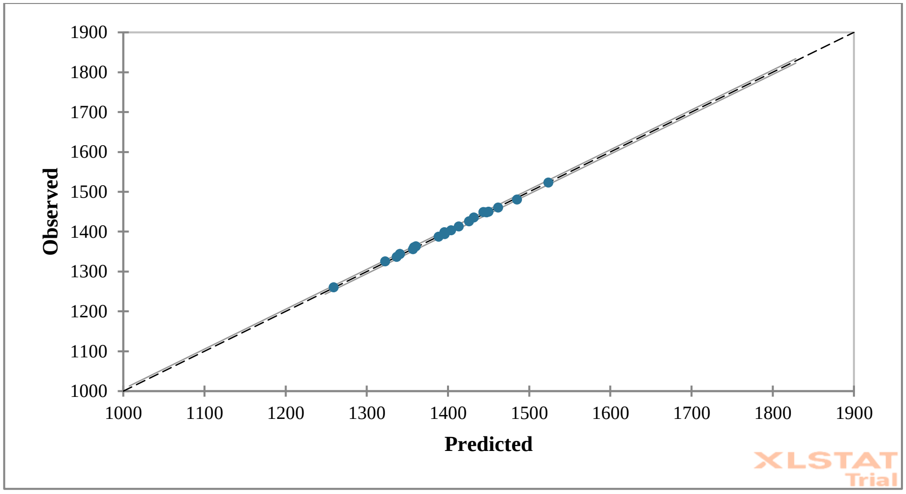

Figure 7: Prediction of multiple linear regression model

Table 5: Prediction of crop yields of 20 plots in sample fields by extreme gradient boosting (XGBoost)

Plot	Observed Yield (g.)	Predicted Yield (g.)	Residual (g.)
1	1,412.480	1,413.179	-0.699
2	1,343.643	1,343.617	0.026
3	1,449.320	1,449.510	-0.190
4	1,259.920	1,259.549	0.371
5	1,324.450	1,324.455	-0.005
6	1,360.260	1,360.715	-0.455
7	1,362.200	1,361.938	0.262
8	1,434.660	1,435.414	-0.754
9	1,336.230	1,336.191	0.039
10	1,386.880	1,387.183	-0.303
11	1,355.260	1,355.333	-0.073
12	1,459.950	1,460.308	-0.358
13	1,397.700	1,398.150	-0.450
14	1,402.782	1,403.048	-0.266
15	1,522.840	1,520.891	1.949
16	1,480.020	1,479.911	0.109
17	1,393.400	1,393.406	-0.006
18	1,447.740	1,447.816	-0.076
19	1,448.440	1,447.990	0.450
20	1,425.520	1,425.089	0.431

Lastly, the accuracy of the results was further analyzed by Extreme Gradient Boosting (XGBoost). The analysis results showed that the prediction of the Extreme Gradient Boosting (XGBoost) model was valid (R^2 = 0.99, RMSE = 0.557g.; MAE = 0.364), as displayed in Figure 8 and Table 5, respectively. According to the results of the MLR model, it was found that the factors affecting the predicted rice yield in the experimental plots, although factors were not positively correlated, were considered from the values of R2, RMSE, and MAE. However, when developed with the XGBoost model, the result was found that the predicted of rice yield from the analyzed factors had a higher level of accuracy. Because Extreme Gradient Boosting (XGBoost) is learning from parameters in the same way as the regression tree family approach. Thus, XGBoost learned from the model parameters and stopped when done to solve the overfitting problem of the results. However, MLR is a statistical method that uses two or more independent variables to estimate the dependent variable only.

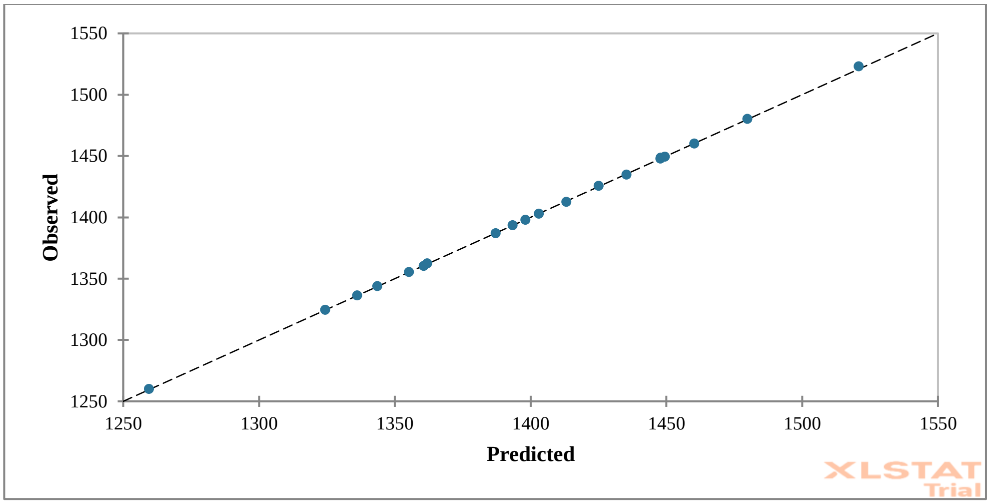

Figure 8: Estimation of extreme gradient boosting (XGBoost) model

4. Conclusion

This research investigates the estimation of crop yields towards the application of unmanned aerial vehicles (UAVs). It was discovered that the UAVs used for estimating crop yields of the sample rice fields in Don Tum District, Nakhon Pathom, could analyze 5 factors affecting the estimation of crop yields, which included three vegetation indexes (Normalized Difference Vegetation Index, Green Normalized Difference Vegetation Index, and Triangular Greenness Index) and two physical covariates (plant height and canopy cover). According to the Multiple Linear Regression analysis, the most significant factors were canopy cover, plant height, and Triangular Greenness Index, followed by Green Normalized Difference Vegetation Index and Normalized Difference Vegetation Index, respectively (R^2 = 0.99). After estimating the crop yields of the sample rice fields based on the five aforementioned factors, the multiple linear regression model showed a positive relationship between the predicted crop yields and the crop yields observed by field exploration (RMSE = 2.506 g.; MAE = 1.778). Additionally, the application of Extreme Gradient Boosting (XGBoost) increased the accuracy of crop yield estimation (RMSE = 0.557 g.; MAE = 0.364). Nonetheless, since this research project was carried out in a genuine agricultural setting, the data collection process was limited by the duration of actual agricultural activities, local climate, and economic conditions of the rice fields' owners. However, the limitation of this study used only factors derived from UAVs. Thus, future studies should include factors related to the environment, economy and farmer behavior that could lead to more accurate yield predictions.

5. Discussion

Since Thailand is one of the world's major rice exporters, rice is a key economic plant of the country. Rice is vastly cultivated nationwide and consumed not only domestically but also internationally [8]. Nowadays, geoinformatics has been widely applied to contemporary agriculture, including rice cultivation, due to its rapid advancement [11]and [12]. Long-distance aerial exploration using optical satellites is frequently used for monitoring rice producing areas and estimating crop yield in time series based on various vegetation indexes [14]. Synthetic Aperture Radar (SAR) satellites are highly suitable for monitoring rainfed rice farming during the rainy season [15]. Apart from satellite imagery or photography, unmanned aerial vehicles (UAVs) have become a popular instrument for modern agriculture as they can provide crucial data for analytical modeling or factor analysis, for example, estimating crop yield by k-means clustering [16]. Therefore, this study employed two statistical tools such as Multiple Linear Regression and Extreme Gradient Boosting (XGBoost) to estimate crop yields. The tools provided valid research results. Furthermore, the factor analysis was performed by two different types of UAVs, including an RGB drone and a multispectral drone [35].

The UAVs discovered that the most significant factors were canopy cover, plant height, and Triangular Greenness Index (TGI), followed by Green Normalized Difference Vegetation (GNDVI) and Normalized Difference Vegetation Index (NDVI), respectively. To increase the accuracy of the plant height estimation, the Light Detection and Ranging (LIDAR) system was also applied [18]. The system is still an expensive technology to date. Even though the UAVs can be used for estimating

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

crop yields, they are high-resolution equipment which is more suitable for exploring small areas. Satellite imagery is thus suggested for monitoring large-scale rice producing areas. Furthermore, Extreme Gradient Boosting (XGBoost) is a model with higher accuracy than MLR considering the results of this study because Extreme Gradient Boosting (XGBoost) is learning from parameters in the same way as the regression tree family approach. However, MLR is a statistical method that uses two or more independent variables to estimate the dependent variable.

Acknowledgement
We would like to express our gratitude to the National Research Council of Thailand (NRCT) for supporting this research. This project is part of Technologies and knowledge transfer to develop rice cultivation potential with Geoinformatics to group of farmers in Nakhonpathom Province. We would like to extend our sincere thanks to Sam Ngam Large-Scale Rice Production Community Enterprise, Don Tum District, Nakhon Pathom, for allowing and facilitating the conduct of field experiments in their property. Lastly, we would like to express our appreciation to the Geography and Geo-Informatics Program, Faculty of Humanities and Social Sciences, Suan Suandha Rajabhat University, for providing supporting equipment and other facilities.

References

[1] Liao, P., Bell, S. M., Chen, L., Huang, S., Wang, H., Miao, J., Qi, Y., Sun, Y., Liao, B., Zeng, Y., Wei, H., Gao, H., Dai, Q. and Zhang, H., (2023). Improving Rice Grain Yield and Reducing Lodging Risk Simultaneously: A Meta-Analysis. *European Journal of Agronomy*, Vol. 143, https://doi.org/10.1016/j.eja.2022.126709.

[2] Prathumchai, K., Nagai, M., N. K. Tripathi. and Sasaki, N., (2018). Forecasting Transplanted Rice Yield at the Farm Scale Using Moderate-Resolution Satellite Imagery and the AquaCrop Model: A Case Study of a Rice Seed Production Community in Thailand, *ISPRS International Journal of Geo-Information*, Vol. 7. https://doi. org/10.3390/ijgi7020073.

[3] Zhang, Z., Li, Y., Xinguo, C., Wang, Y., Niu, B., Liu, D. L., He, J., Pulatov, B., Hassan I., and Meng, Q., (2023). Impact of Climate Change and Planting Date Shifts on Growth and Yields of Double Cropping Rice in Southeastern China in Future. *Agricultural Systems*, Vol. 205. https://doi.org/10.1016/j.agsy.2022.103581.

[4] Arunrat, N., Pumijumnong, N., Sereenonchai, S., Chareonwong, U. and Wang, C., (2020). Assessment of Climate Change Impact on Rice Yield and Water Footprint of Large-Scale and Individual Farming in Thailand. *Science of the Total Environment*, Vol. 726. https://doi.org/ 10.1016/j.scitotenv.2020.137864.

[5] Nyussupova, G., Tokbergenova, A., Aubakirova, G., Kelinbayeva R. and Kalimurzina, A., (2019). The Electronic Atlas of Socio-Demographic Development of the Regions of the Republic of Kazakhstan: Elaboration and Main Stages of Creation. *International Journal of Geoinform atics*, Vol. 15, 67-75.

[6] Liu, X., Zheng, J., Yu, L., Hao, P., Chen, B., Xin, Q., Fu, H. and Gong, P., (2021). Annual Dynamic Dataset of Global Cropping Intensity from 2001 to 2019. *Scientific Data*, Vol. 8. https://doi.org /10.1038/s41597-021-01065-9.

[7] Frazier, A. G., Yen, B. T., Stuecker, M. F., Nelson, K. M., Sander, B. O., Kantar, M. B. and Wang, D. R., (2022). Impact of Historical Climate Variability on Rice Production in Mainland Southeast Asia Across Multiple Scales, *Anthropocene*, Vol. 40. https://doi.org/ 10.1016/j. ancene.2022.100353.

[8] Silva, J. V., Pede, V. O., Radanielson, A. M., Kodama, W., Duarte, A., de Guia, A. H., Malabayabas, A. J. B., Argosubekti, N., Vithoonjit, D., Hieu, P. T. M., Pame, A. R. P., Singleton, G. R. and Stuart, A. M., (2022). Revisiting Yield Gaps and the Scope for Sustainable Intensification for Irrigated Lowland Rice in Southeast Asia. *Agricultural Systems*, Vol. 198. https://doi.org/10.1016/j.agsy.2022. 103383.

[9] Methamontri, Y., Tsusaka, T. W., Zulfiqar, F., Yukongdi, V. and Datta, A., (2022). Factors Influencing Participation in Collective Marketing through Organic Rice Farmer Groups in Northeast Thailand. *Heliyon*, Vol. 8. https://doi. org/10.1016/j.heliyon.2022.e11421.

[10] Office of Agricultural Economics, (2022). Agricultural Economic Data, [Online]. Available: Office of Agricultural Economics, https://www. oae.go.th. [Accessed December 31, 2022].

[11] Zhou, X., Zheng, H., Xu, X., He, J., Ge, X., Yao, X., Cheng, T., Zhu, Y., Cao, W. and Tian, Y., (2017). Predicting Grain Yield in Rice Using Multi-Temporal Vegetation Indices from UAV-Based Multispectral and Digital Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 130, 246-255. https://doi.org/10. 1016/j.isprs jprs.2017.05.003.

[12] Shen, Y., Mercatoris, B., Cao, Z., Kwan, P., Guo, L., Yao, H. and Cheng, Q., (2022). Improving Wheat Yield Prediction Accuracy Using LSTM-RF Framework Based on UAV Thermal Infrared and Multispectral Imagery. *Agriculture*, Vol. 12. https://doi.org/10.3390/agriculture12060892.

[13] Worachairungreung, M., Thanakunwutthirot, K. and Kulpanich, N., (2023). A Study on Oil Palm Classification for Ranong Province Using Data Fusion and Machine Learning Algorithms. *Geographia Technica.*, Vol.18, 161-176. http://doi.org/10.21163/GT_2023.181.12.

[14] González, J. S., Angelats, E., Eixarch, M. M. and Alcaraz, C., (2022) Monitoring Rice Crop and Yield Estimation with Sentinel-2 Data, *Field Crops Research*. Vol. 281, https://doi.org/10.1016/j.fcr.2022.108507.

[15] Hashemi, M. G., Abhishek, A., Jalilvand, E., Jayasinghe, S., Andreadis, K.M., Siqueira, P. and Das, N. N., (2022). Assessing the Impact of Sentinel-1 Derived Planting Dates on Rice Crop Yield Modeling, *International Journal of Applied Earth Observation and Geoinformation*, Vol 114. https://doi.org/10.1016/j.jag.2022.103047.

[16] Reza, M. N., Na, I. S., Baek, S. W. and Lee, K. H., (2019). Rice Yield Estimation Based on K-means Clustering with Graph-Cut Segmentation Using Low-Altitude UAV Images, *Biosystems Engineering*, Vol. 177, 109-121. https://doi.org/10.1016/j.biosystemseng.2018.09.014.

[17] Ge, H., Ma, F., Li, Z. and Du, C., (2022). Estimating Rice Yield by Assimilating UAV-Derived Plant Nitrogen Concentration into the DSSAT Model: Evaluation at Different Assimilation Time Windows, *Field Crops Research*, Vol. 288. https://doi.org/10.1016/j.fcr.2022.108705.

[18] Phan, A. T. and Takahashi, K., (2021). Estimation of Rice Plant Height from a Low-Cost UAV- Based LIDAR Point Clouds. *International Journal of Geoinformatics*, Vol. 17, 89–98. https://doi.org/10.52939/ijg.v17i2.1765.

[19] Kulpanich, N., Worachairungreung, M. and Chaysmithikul, P., (2020). Surveying for Built-Up Areas in Nakhon Pathom Province with Remote Sensing from Landsat 8 and Sentinel 2 Imagery. *Suan Sunandha Academic & Research Review*, Vol. 14, 44-55.

[20] Land Development Department, (2022). Soil Service - Land Use.

[21] DJI, (2018). Phantom 4_Pro and Pro Plus Series User Manual.

[22]Kulpanich, N., Worachairungreung, M., Waiyasusri, K., Saengow, P. and Pinasu, D., (2022). Height Measurement and Oil Palm Yield Prediction Using Unmanned Aerial Vehicle (UAV) Data to Create Canopy Height Model (CHM). *Geographia Technica*, Vol. 17, 164-178. https://doi.org/10.21163/GT_2022.172.14.

[23] DJI, (2020). P4 Multispectral User Manual v1.4 EN.

[24] Kong, J., Ryu, Y., Huang, Y., Dechant, B., Houborg, R., Guan, K. and Zhu, X. (2021). Evaluation of Four Image Fusion NDVI Products Against In-Situ Spectral-Measurements Over a Heterogeneous Rice Paddy Landscape. *Agricultural and Forest Meteorology*, Vol. 297. https://doi.org/10.1016/j.agrformet.2020.108255.

[25] Basso, M., Stocchero, D., Henriques, R. V. B., Vian, A. L., Bredemeier, C., Konzen, A. A. and de Freitas, E. P., (2019). Proposal for an Embedded System Architecture Using a GNDVI Algorithm to Support UAV-Based Agrochemical Spraying. *Sensors*, Vol. 19. https://doi.org/10.3390/s19245397.

[26] Hunt, Jr. E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S., Perry, E. M. and Akhmedov, B., (2013). A Visible Band Index for Remote Sensing Leaf Chlorophyll Content at the Canopy Scale. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 21, 103-112. https://doi.org/10.1016/j.jag.2012.07.020.

[27] Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., Du, X., Zhai, L., Weng, H., Li, Y., Li, X., Bao, Y., Shou, J. and He, Y., (2020). Grain Yield Prediction of Rice Using Multi-Temporal UAV-Based RGB and Multispectral Images and Model Transfer – A Case Study of Small Farmlands in the South of China. *Agricultural and Forest Meteorology*, Vol. 291. https://doi.org/10.1016/j.agrformet.2020.108096.

[28] Malambo, L., Popescu, S., Murray, S., Putman, E., Pugh, N., Horne, D., Richardson, G., Sheridan, R., Rooney, W., Avant, R., Vidrine, M., McCutchen, B., Baltensperger, D. and Bishop, M., (2018). Multitemporal Field-Based Plant Height Estimation Using 3D Point Clouds Generated from Small Unmanned Aerial Systems High-Resolution Imagery. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 64, 31-42. https://doi.org/10.1016/j.jag.2017.08.014.

[29] Hunt, E. R., Hively, W. D., McCarty, G. W., Daughtry, C. S. T., Forrestal, P. J., Kratochvil, R. J., Carr, J. L., Allen, N. F., Fox-Rabinovitz, J. R. and Miller, C. D., (2011). NIR-Green-Blue High-Resolution Digital Images for Assessment of Winter Cover Crop Biomass. *GIScience & Remote Sensing*, Vol. 48, 86-98. https://doi.org /10.2747/1548-1603.48.1.86.

[30] Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R. and Chehaibi, S., (2019). Prediction of Organic Potato Yield Using Tillage Systems and Soil Properties by Artificial Neural Network (ANN) and Multiple Linear Regressions (MLR). *Soil and Tillage Research*, Vol. 190, 202-208. https://doi.org/10. 1016/j.still.2019.01.011.

[31] Wang, J., Wu, B., Kohnen, M., Lin, D., Yang, C., Wang, X., Qiang, A., Liu, W., Kang, J., Li, H., Shen, J., Yao, T., Su, J., Li, B. and Gu, L., (2021). Classification of Rice Yield Using UAV-Based Hyperspectral Imagery and Lodging Feature. *Plant Phenomics*, Vol. 2021, 1-14. https://doi.org /10.34133/2021/9765952

[32] Li, P., Hao, H., Bai, Y., Li, Y., Mao, X., Xu, J., Liu, M., Lv, Y., Chen, W. and Ge, D., (2022). Convolutional Neural Networks-Based Health Risk Modelling of Some Heavy Metals in a Soil-Rice System. *Science of the Total Environment*, Vol. 838. https://doi.org/10.1016/j.scitotenv. 2022.156466.

[33] Chu, Z. and Yu, J., (2020). An End-to-End Model for Rice Yield Prediction Using Deep Learning Fusion. *Computers and Electronics in Agriculture*, Vol. 174. https://doi.org/10.1016/ j.compag.2020.105471.

[34] Abdikan, S., Sekertekin, A., Narin, O. G., Delen, A. and Sanli, F. B., (2022). A Comparative Analysis of SLR, MLR, ANN, XGBoost and CNN for Crop Height Estimation of Sunflower Using Sentinel-1 and Sentinel-2. *Advances in Space Research*, Vol. 71(7), 3045-3059. https://doi.org/10.1016/j.asr.2022.11.046.

[35] Bascon, M. V., Nakata, T., Shibata, S., Takata, I., Kobayashi, N., Kato, Y., Inoue, S., Doi, K., Murase, J. and Nishiuchi, S., (2022). Estimating Yield-Related Traits Using UAV-Derived Multispectral Images to Improve Rice Grain Yield Prediction. *Agriculture*, Vol. 12(8). https://doi.org/10.3390/agriculture12081141.