Rice Yield Estimation Based on Machine Learning Approaches using MODIS 250 m Data

Woranan Mongkolnithithada [1], Jurawan Nontapun [1], and Siwa Kaewplang [1*]

[1] Faculty of Engineering, Mahasarakham University, Tumbol Khamriang, Ampher Kantarawichai, Mahasarakham, 44150, Thailand

vplove888@gmail.com, jurawannontapon@gmail.com and siwa.kae@msu.ac.th* (corresponding author)

Abstract. *Food security and water resource management depend on knowledge about the distribution of paddy rice fields. Information on rice production from space can be obtained using the technology of remote sensing. In the current study, the relationships between the rice spectrum, vegetation index, and rice yield can be assessed using the Moderate Resolution Imaging Spectroradiometer (MODIS). Machine learning has been evolving steadily in recent years, and its benefits are now readily apparent; particularly in the area of image processing, it is advancing quickly. The objective of this research is to estimate the rice yield using the MODIS satellite imagery data based on machine learning. Three machine-learning regression algorithms (multiple linear regression, support vector machine, and random forest) were evaluated, and a suitable model was created to estimate the rice production. According to the findings, the random forest model produced the most objective results, had the lowest RMSE values, and had good statistical correlations for both the training set and the test set. The methods described in this paper can be used as a reference for combining machine learning with MODIS satellite imagery data to estimate the rice yield in other locations.*

Received by	05 April 2023
Revised by	18 April 2023
Accepted by	19 April 2023

Keywords:

Machine learning, Rice yield, Modis 250 m, Remote sensing

1. Introduction

In Thailand, rice is one of the most significant agricultural crops. Globally, rice is grown on around 160.5 million hectares of land, comprising about 11.5% of all arable land. In a world where production is frequently insufficient to meet demand due to the growing population, rice area mapping and production forecasts are crucial to food security [1].

Satellite remote sensing has been extensively applied and is acknowledged to be a powerful and useful technology for identifying changes in land use and land cover [2-4]. Research using satellite imaging has been carried out to monitor the growth and yields of rice [5-8].

The estimation of rice yield from remote sensing data has been achieved with Landsat images [9-12]. However, the use of this type of estimation is constrained by the lack of data for temporal analysis due to the 16-day temporal resolution of Landsat [12]. Spatial resolution is an additional constraint on how NOAA images can be used. The probability of a single pixel comprising multiple types of land cover is high, especially in small study areas, because one pixel in NOAA images may correspond to 1,000 m [13]. The accuracy of assessment is lowered by the existence of various land uses in a single pixel [14-15]. The Moderate Resolution Imaging Spectroradiometer (MODIS), however, with 250 m spatial resolution, offers superior spatial resolution as compared to NOAA images, and better temporal resolution than Landsat images [14-15]. Moreover, the Normalized Difference Vegetation Indices (NDVI), which are based on reflectance in the red and near infrared regions, and which are obtained from multi-temporal satellite images, provide excellent indicators of the capacity and efficiency of plant photosynthetic processes, and can be used to predict crop yield [16-19].

The highest biomass at the heading stage is highly correlated with rice crop production [20], Previous research has revealed that the ripening stage has the strongest relationship coefficients with crop yields and vegetation indices, and the grain filling period [21-22]. Thus, we hypothesized that the Modis Data surrounding the heading date would be significantly correlated with yields of rice crops.

Machine learning is an automated, data-driven, self-adaptive process that does not require awareness of the physical relationships or systems that generate the data [23]. The machine is given access to data so it may self-validate and learn. It develops information about the functional relationships through training, and predictions are validated.

For this reason, machine learning techniques provide extremely accurate and simple ways to identify patterns and rules within big data sets containing numerous predictor-variables that have nonlinear relationships with the target variable.

It has been shown that remotely sensed data and machine learning can be effectively used in estimating the crop yield [24-28]. For example, monitoring of mineral nutrition of arabica coffee crops [24], and evaluation of wheat yields are carried out with data from simulation modeling and remote sensing-based machine learning techniques [25]. Sentinel-2 satellite data and deep machine learning are used to estimate paddy rice production [26], and machine learning is utilized for rice crop yield predictions using Modis data [27-28].

The main objective of this study was to develop an approach to bridge an existing research gap between the effectiveness of MODIS data and machine learning for a large-scale yield estimation of rice crops in Ubonrachatani Province, Thailand. The hypothesis predicts a significant correlation with rice crop yeilds.

2. Materials and Methods

2.1 Study Area

The study area focuses on a rice plantation in Ubonrachatani Province Thailand, that cover approximately 500 km2 at 15°20'56.75" N latitude and 104°50'31.60" E longitude as shown in the yellow layer in Figure 1. The topographic elevation generally ranges between 100 and 400 meters above mean sea level. The mean annual rainfall of Ubonrachatani Province is 1,270 mm. The rainy season, which sets in during 2nd week of May and continues till the 2nd week of October, has a mean seasonal rainfall of 1,050 mm.

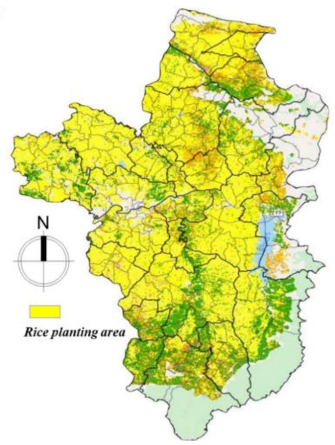

Fig. 1 Map of the study area: Ubonrachatani Province, Thailand

2.2 Rice Grain Sample Collection

The following factors were considered in selection of the rice grain sampling sites: (1) the planted area should be wide and easily identifiable in MODIS images; and (2) the planted area should comprise a single rice variety. At each sampling site, rice grain was collected from a 25 m x 25 m area. The weight of the rice grain was measured in kg units, and the yield was obtained in tons/ha.

2.3 MODIS Images

We used the 8-day composite MODIS Surface Reflectance Band 1-2 Product, one of a number of standard MODIS products available to users (MYD09Q1), with data gathered from the 2018 and 2019 rice cropping seasons (Oct 2018– Oct 2019) for use in this study. This MODIS product has a spatial resolution of 250 m and two bands, including the red band (620-670 nm) and the infrared band (841–876 nm). Additionally, atmospheric corrections for gases, thin cirrus clouds, and aerosols are implemented in the production of MOD09A1. We utilized the Modis images when the rice plant cycle was anticipated to reach NDVI maximum (as shown in Figure 2). The MODIS images were downloaded at no charge from the MODIS website. (http://mrtweb.cr.usgs.gov/) utilizing the USGS MODIS Reprojection Tool Web Interface (MRTWeb).

2.4 . Data Analysis

The imagery was pre-processed; rice spectral data was gathered from the MODIS images, and statistical analyses were carried out throughout the data processing. The website's MODIS imagery was downloaded; cloud-free images were chosen and cropped, and the geographic locations were changed in order to pre-process the data. The correlation between rice yield and MODIS Reflectance was discovered using the red and infrared bands from MODIS images that were collected from different places.

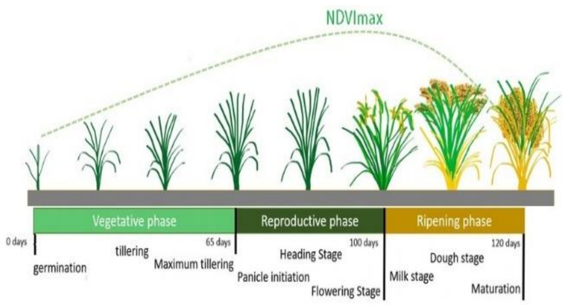

Fig. 2 Rice plant cycle with phase and froth stages; the curve represents the position profile for NDVImax [29].

3. Modeling and Re-sampling

Three models were created in an attmpt to identify the best one for using reflectance values to evaluate estimates of rice yield. , Multiple linear regression, support vector machine, and random forest models are strict in requiring predictor variables with a common scale, ; therefore, pre-processing of data must be done on the training set before modeling. The coefficient of determination (R^2), and RMSE were utilized as measures to evaluate the regression model's performance and evaluate the its complexity as well with respect to how effectively it performs with new data. Equations (1) and (2) are used to calculate R^2, and RMSE:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \quad 1)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad 2)$$

Where N is the total sample size , ; y_i is the ith measured biomass of the sample, ; $y^\wedge_i$ is the ith predicted value, and $\bar{y}_i$ is the ith mean measured value.

4. Results and Discussion

The field-measured average rice yield value was 2800 kg/ha (N = 400, SD = 855). Table 1 shows the R^2 and RMSE values for the multiple linear regression, random forest and support vector machine models. The greatest value is underlined. Figure 3 displays the plots for each model. The comparability of the three different regression models was also examined using a one-way ANOVA test. It turns out that there is no statistically significant difference between the three models (i.e., p-value<0.01, N = 400). In other words, all three machine-learning regression approaches are correlated to the field rice yield. (The multiple linear regression, random forest, and support vector machine models) derived from reflectance values yielded R^2 = 0.537, 0.644, and 0.760, and RMSE = 360.9, 350.8, and 291.7, respectively. The Support Vector Machine's R2 value was high, proving the consistency of the relationship between rice yield and the Modis Data. Moreover, the RMSE value was low, indicating there was a small difference between the estimated result and the actual value [27-28].

Furthermore, the near-infrared (NIR) band of the rice plant was high. Green biomass and rice NIR reflectance are closely connected [16-19]. Thus, the high reflectance of the near-infrared (NIR) band and low reflectance of the red band indicate high NDVI values, and are indicative of high chlorophyll content. Chlorophyll is the most important part of the rice plant for photosynthetic activity as it produces carbohydrates to form rice plant tissue and rice grains, which has a significant effect on the yield at harvest [20-22]. This study demonstrates a strong relationship between MODIS data and rice yield, with [27-28] describing similar results.

There could be several reasons for the difference between rice yields in actual and predicted conditions, such as cloud interference in the atmosphere [30], weather variations on a micro level, or water availability [31] as well as uncertainty related to ground-based estimations [32]. The R2 and RMSE of the MODIS data are significantly impacted by the presence of clouds. The reflectance values may vary as a result of thin cloud cover, which will have an impact on the MODIS data for the rice field. As a result, one of the most important steps in data processing is selecting imagery without clouds. Moreover, mixed pixels can degrade the rice spectral information and impact the R2 and RMSE of the MODIS data.

Model	Training		Testing	
	R^2	RMSE	R^2	RMSE
Multiple linear regression	0.479	399.3	0.537	360.9
Support vector machine	0.611	377.2	0.644	350.8
Random forest	0.683	343.9	***0.760***	***291.7***

Table 1 The R^2 values and the RMSE values of three machine-learning regression algorithms indicating P<0.01

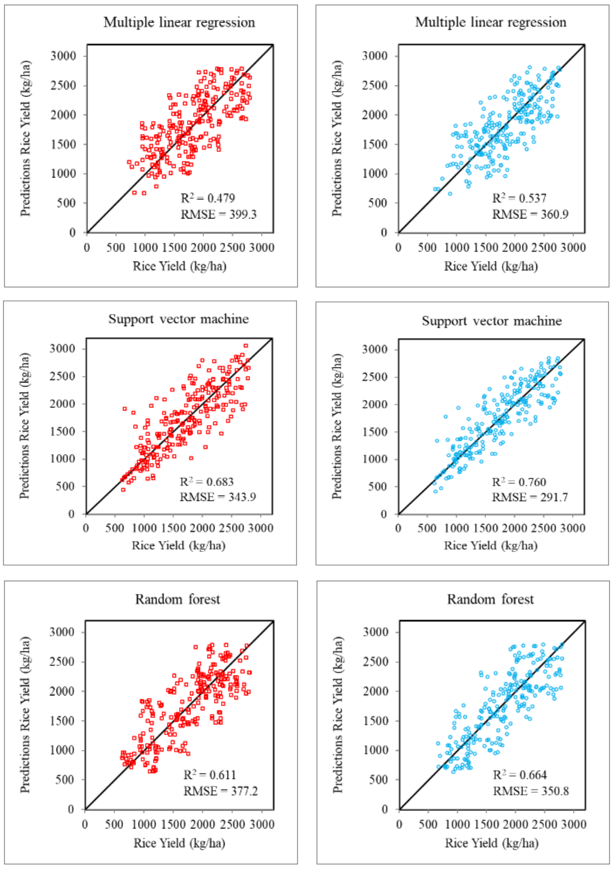

Fig. 3 The scattering plots with the RMSE values of all models

5. Conclusion

The findings of this study support the use of MODIS 250 m data for rice yield estimation as a first step in conducting more research. The three machine- learning regression approaches are each correlated with the field rice yield. (Multiple linear regression, random forest, and support vector machine models) derived from reflectance values yielded $R^2 = 0.537$, 0.644, and 0.760, and RMSE = 360.9, 350.8, and 291.7, respectively. The support vector machine model displayed the lowest error rate when compared to independent field data with strong statistical correlations of the proposed rice yield models.

In conclusion, this study takes the previous research a step further. by investigating the potential utility of MODIS 250 m data for estimating rice yield with machine learning. The proposed statistical models, despite the need for further fine-tuning of the model parameters, exhibit high statistical correlations (best R2 = 0.760) with low RMSE values (lowest RMSE = 291.7). Therefore, we assume that the steps outlined in this study will be helpful in guiding rice yield determination for other areas.

Acknowledgements

This study was supported by, Faculty of Engineering, Mahasarakham University, Thailand. The financial support from Medium Irrigation Construction Office 7 Medium Water Resources Development Division, Royal Irrigation Department, Thailand.

References

[1] PRATHUMCHAI, Kulapramote, 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*, 2018, pp.72-73.

[2] JIANG, Jing; TIAN, Guangjin. Analysis of the impact of land use/land cover change on land surface temperature with remote sensing. *Procedia environmental sciences*, 2010, pp.571-575.

[3] REIS, Selçuk. Analyzing land use/land cover changes using remote sensing and GIS in Rize, North-East Turkey. *Sensors*, 2008, pp.6188-6202.

[4] KAFY, Abdulla-Al, et al. Remote sensing approach to simulate the land use/land cover and seasonal land surface temperature change using machine learning algorithms in a fastest-growing megacity of Bangladesh. *Remote Sensing Applications: Society and Environment*, vol. 21, pp. 100463, 2021, doi:10.1016/j.envc.2021.100190.

[5] ZHU, Qiqi, et al. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, pp. 63-78

[6] FROLKING, Steve, et al. Agricultural land-use in China: a comparison of area estimates from ground-based census and satellite-borne remote sensing: GCTE/LUCC RESEARCH LETTER. *Global Ecology and Biogeography*, 1999, pp.407-416.

[7] HÜTT, Christoph, et al. Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images. *Remote sensing*, vol.8, pp. 684, 2016, doi:10.3390/rs8080684.

[8] FROLKING, Steve, et al. Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China. *Global Biogeochemical Cycles*, 2002, pp. 38-1-38-10.

[9] NUARSA, I. Wayan; NISHIO, Fumihiko; HONGO, Chiharu. Rice yield estimation using Landsat ETM+ data and field observation. *Journal of Agricultural Science*, 2012, pp.43-45.

[10] SIYAL, Altaf Ali; DEMPEWOLF, Jan; BECKER-RESHEF, Inbal. Rice yield estimation using Landsat ETM+ Data. *Journal of Applied Remote Sensing*, vol.9, pp.095986-095986, 2015, doi:10.1117/1.JRS.9.095986.

[11] SHEW, Aaron M.; GHOSH, Aniruddha. Identifying dry-season rice-planting patterns in bangladesh using the landsat archive. *Remote Sensing*, vol.11, pp.1235, 2019, doi:10.3390/rs11101235.

[12] SAADI, Terry Devara Tri; WIJAYANTO, Arie Wahyu. Machine learning applied to sentinel-2 and landsat-8 multispectral and medium-resolution satellite imagery for the detection of rice production areas in Nganjuk, East Java, Indonesia. *International Journal of Remote Sensing and Earth Sciences*, 2021, pp.19-32.

[13] FANG, Hongliang, et al. Using NOAA AVHRR and Landsat TM to estimate rice area year-by-year. *International Journal of Remote Sensing*, 1998, pp.521-525.

[14] NUARSA, I. Wayan; NISHIO, Fumihiko; HONGO, Chiharu. Relationship between rice spectral and rice yield using MODIS data. *Journal of Agricultural Science*, vol.3, pp.80-88, 2011, doi:10.5539/jas.v3n2p80.

[15] SHIHUA, Li, et al. Monitoring paddy rice phenology using time series MODIS data over Jiangxi Province, China. *International Journal of Agricultural and Biological Engineering*, 2014, pp.28-36.

[16] HUANG, Jingfeng, Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. *PloS one*, vol.8, pp. e70816, 2013, doi : 10.1371/journal.pone.0070816.

[17] ZHOU, Yuting, "Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images". *International Journal of Applied Earth Observation and Geoinformation*, 2016, pp. 1-12.

[18] ZHANG, Meng, "Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. *Remote Sensing*, vol. 10, pp. 1840, 2018, doi:10.3390/rs10111840.

[19] QIN, Yuanwei, et al. Mapping paddy rice planting area in cold temperate climate region through analysis of time series Landsat 8 (OLI), Landsat 7 (ETM+) and MODIS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, pp. 220-233.

[20] HONDA, Sotaro, Maintaining higher leaf photosynthesis after heading stage could promote biomass accumulation in rice. *Scientific reports*, vol.11, pp.7579, 2021, doi:10.1038/s41598-021-86983-9.

[21] PANUJU, Dyah R., Mapping rice growth stages employing MODIS NDVI and ALOS AVNIR-2. *Remote Sensing and GIScience: Challenges and Future Directions*, 2021, pp.185-203.

[22] ISLAM, Md Monirul, et al. Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. *Remote Sensing Applications: Society and Environment*, vol.22, pp.100494, 2021, doi:10.1016/j.rsase.2021.100494.

[23] ZHOU, Lina, et al. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 2017, pp.350-361.

[24] DE CARVALHO ALVES, Marcelo, et al. The role of machine learning on Arabica coffee crop yield based on remote sensing and mineral nutrition monitoring. *Biosystems Engineering*, 2022, pp. 81-104.

[25] MERAJ, Gowhar, et al. Assessing the yield of wheat using satellite remote sensing-based machine learning algorithms and simulation modeling. *Remote Sensing*, vol. 14, pp. 3005, 2022, doi:10.3390/rs14133005.

[26] THORP, K. R.; DRAJAT, D. E. N. A. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia. *Remote Sensing of Environment*, vol.265, pp.112679, 2021, doi: https://doi.org/10.1016/j.rse.2021.112679.

[27] MA, Jong-Won, et al. Regional-scale rice-yield estimation using stacked auto-encoder with climatic and MODIS data: A case study of South Korea. *International Journal of Remote Sensing*, 2019, pp.51-71.

[28] SON, Nguyen-Thanh, et al. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. *International Journal of Remote Sensing*, 2020, pp.7868-7888.

[29] Mosleh, M. K., Hassan, Q. K., and Chowdhury, E. H. (2015). Application of remote sensors in mapping rice area and forecasting its production: A review. Sensors (Switzerland), pp.769–791.

[30] Mkhabela, M.S., Bullock P., Raj S., Wang S., andYang Y. 2011. Crop yield forecasting on the Canadian prairies using MODIS NDVI data. Agric. For. Meteo.,2011. pp.385-393.

[31] Son, N. T., Chen, C. F., Chen, C.R., Chang, L.Y., Duc, H.N., Nguyen, L.D., 2013. Prediction of rice crop yield using MODIS EVI-LAI - data in the Mekong Delta, Vietnam. Int. J. Remote Sens. ,2013. pp.7275-7292.

[32] Mosleh, M.K., Hassan, Q.K., 2014. Development of a remote sensing-based "Boro" rice mapping system. Remote Sens.,2014. pp. 1938-1953

Biographies

Woranan Mongkolnithithada received Her bachelor degree in civil engineeringfrom Western University, Thailand, in 2013. She is currently a Irrigation Technician, Royal Irrigation Department, Thailand. her current Research interests include RemoteSensing ;Image classification; Image processing; Information technology; Precision agriculture**.**

Jurawan Nontapon is a Ph.D. Candidate in civil engineering from Mahasarakham University, Thailand. His research interests include Digital Photogrammetry; Geographic information systems (GIS); remote sensing; image classification; image processing; information technology; precision agriculture; Water Resource Model;Water Resource Management; Climate Chang; Saline Soil

Siwa Kaewplang received his Phd. from Chulalongkorn University, Thailand, in 2014. He is currently an Assistance Professor of civil engineering at the Faculty of Engineering, Mahasarakham University, Thailand. His research interests include Digital Photogrammetry; Geographic Information System; Geostatistics; Remote sensing; Hyperspectral remote sensing; Image classification; Image processing; Information technology; Precision agriculture; Natural resource management; Tropical vegetation; Water resource management and irrigation engineering.