Paddy stage identification and paddy yield forecasting using satellite-based technology

Abstract — **Rice is the main food of the inhabitants of Sri Lanka. Hence, monitoring at every growth phase of a rice plant is important information fact for an estimate the yield prediction before harvest. It helps many aspects like strategic planning and decision making regarding the food security and facilitation of safe harvest storages. In Sri Lanka, a traditional "crop cutting survey" fails to forecast rice yield before harvest as this experiment is conducted during the harvest. However, previous research studies have concluded that satellite RS (Remote Sensing) as a potential tool for identifying the paddy stages and estimating and forecasting crop yields. In this study, used Sentinel-2A free RS data from 2015 to 2019 over Gampaha, Kiridiwita selected paddy area, for classifying the paddy stages (Germination, Vegetative phase, Reproductive phase and Ripening phase) and also build the model for selected area harvest predicting. For classifying the paddy stages, used image classification using Support Vector Machine(SVM) approach. Since Sentinel-2A carries the Multispectral Imager (MSI) then (B04, B03, B02), (B11, B08, B04), (B11, B08, B02) bands combinations are selected to build the image classification model for identifying the paddy crop stages. Among them (B11, B08, B02) bands combination shows a more accurate model with an average overall accuracy of 92%. For forecasting the paddy yield of the selected area, analysis the linear gradient and polynomial coefficients in order two of NDVI(Normalized Difference Vegetation Index) values of every Yala, Maha season in each year and map with the past harvest data and build the model. The gradient of the linear in temporal NDVI paddy seasonal data(Yala & Maha) vs past harvest data shows more correlation with $\rho = 0.95$ other than Polynomial coefficients. So paddy yield forecasting selected model was $y = 11.95704437m + 1.29989691$ in AOI. Once the paddy crop is identified as in Ripening phase, it shows the paddy yield forecasting result per acre. Therefore, this experiment has demonstrated that the seasonal rice yield can successfully be forecasted one month prior to harvest with considerably higher accuracy.**

Keywords — paddy yield forecasting, paddy stage identification, Sentinel-2 images, NDVI, SVM

I. INTRODUCTION

Rice is the main food of the inhabitants of Sri Lanka. Paddy crops are cultivated as a wetland crop in all the districts. The total land devoted to paddy is estimated to be about 708,000 Hectares at present [1]. There are two cultivation seasons namely Maha and Yala which are synonymous with two monsoons in Sri Lanka. According to Paddy Statistics of Department of the census and Statistic, the average yield of paddy estimated for 2017 Yala season was 83.2 bushels per net acre. It decreases 2.4 bushels per net acre compare to 2016 Yala Season and Paddy estimated 2017/2018 Maha season was 83.43 bushels per net acre. It increases 123.26 bushels per acre compare to previous Maha season. The aforementioned data infers the growth of the rice yield per unit area, and also reflects the development of the national economy. Therefore, precisely mapping the cultivation is and estimating the yield of paddy rice are both crucial for national food security and national development evaluation.

In Sri Lanka, Average paddy yield is estimated by a simple survey. This process, commonly known as "crop cutting survey", was introduced by the Food and Agriculture Organization of the United Nations (FAO) in 1951. At present, a sample of 3,000 villages for the main season (Maha) and 2,000 villages for the second season (Yala) are selected to carry out experiments. [1]. However, this method fails to predict rice yield before the harvest as the sampling is conducted during the harvest. As well as, this survey is usually time-consuming, subjective, costly and likely to significant errors because of the limited ground survey, resulting in poor rice yield estimations.

Paddy yield forecast a few months before the harvest can be of huge importance to take necessary management decision regarding seasonal crop productions. In addition, it is most important to strategic planning and decision making regarding the food security and facilitation of safe harvest storages. According to the Sri Lankan Department of Agriculture, in the 2015/2016 Maha and 2016 Yala seasons there were 2.5 and 1.3 million metric tons of rice produced, respectively. However, during these two seasons, it was recorded that farmers in many parts of the country waited days near the storage facilities to receive payment for their crop. In the end, they were disappointed over the failure to set up an adequate mechanism to purchase their harvest. On the other hand, in early 2017, the price of rice increased significantly due to the inadequate stores of rice within the country. These two recent incidents reflect the inadequate management and poor decision making of national rice production in Sri Lanka. The conditions of surplus in 2015/2016 and scarcity in 2017 were primarily due to the unavailability of a system to forecast the expected paddy yield by a considerable time before harvesting [2].

So, to overcome previously mentioned cases by proper import, export or rearrange storage facilities, taken based on reliable yield estimations as suggested by N.A. Noureldin [3]. Recently, there are lunching various satellites like Landsat, Sentinel, Tera etc with particular remote sensors to crop monitoring. Whereas, remote sensing data have the potential to provide timely, systematic high quality spatial and accurate details about land feature including the environmental impact of the crop growth [4]. Using sentinel-2A satellite data, creating a machine learning model to identify the stage of the paddy crop and paddy yield prediction in the ripening phase in real-time.

This research addresses this very important and urgent issue on how to predict the harvest through NDVI values in the third phase(ripening phase) of paddy rice using remote sensing images with support vector machines. The non-destructive nature and capabilities of large scale observation and on-site monitoring makes remote sensing technique being an ideal means used for implementation practices in rice production [5]. [6] proposed a rice yield prediction method based on land and airborne hyperspectral images. It involved laborious and high-cost campaigns both infield and airborne. It was aimed as a capacity building measure to anticipate the launching of the next generation of Hyperspectral spaceborne satellite (Hyperspectral Image Suite: HISUI). In the absence of hyperspectral satellite imagery, it is very difficult to extrapolate the application of our model and thus to extend our area of estimations. Whereas, the airborne campaign is quite expensive. Instead of waiting and relying on the availability of hyperspectral satellite images, we study the use of Sentinel-2A which is a kind of multispectral remote sensing satellite operated and developed by the ESA(European Space Agency). [7] deduce with their research that Satellite Remote Sensing (RS) can be successfully applied for this particular research area. The most multispectral satellite system measure varies spectral bands including the visible to the mid-infrared region of the electromagnetic spectrum [8]. The main advantage of sentinel-2A is its direct broadcast capability, and anyone can constantly download this broadcast data free of charge. Moreover, its acquiring system almost covers all of the earth surfaces, and acquire every 4-5 days at the same place, so make it possible to apply for monitoring an object on the earth surface. It plays a vital role in broad applications such as the development of the prediction model for crop yields and classification of crops using multidimensional regression of NDVI and surface temperature [9].

The analysis of Sentinel-2A data for paddy growth stages classification is not a trivial task due to many factors such as large spatial variability of each growth stage due to non-uniform plantation time, atmospheric effects and the curse of dimension in dealing with a larger number of frequency bands. [10] develops a novel approach to analyze satellite remote sensing images, particularly Sentinel-2A satellite images using machine learning techniques. [11] address the problem of harvest time and area prediction through growth stages classification using MODIS remote sensing data with support vector machines.

Regarding with yield prediction model literature, [12] examines the influences of cirrus cloud on spectral indices of vegetation like NDVI and AFRI. The spectral absorption typically occurs from 670 to 780 nm wavelength range of the electromagnetic spectrum [13]. In this case, leaf chlorophyll has a strong absorption at 0.45 μm and 0.67 μm, and high reflectance at near-infrared (0.7–0.9 μm). Healthy plants have a high NDVI (normalized difference vegetation index) value because of their high reflectance of infrared light, and relatively low reflectance of the red spectrum [14]. The model is based on vegetation indices (VI) which could be collected from remote sensing satellite data. VI are optical measures of vegetation canopy ‘‘greenness’’. Some researches have observed that there is a linear relationship between rice yield and VI calculated based on satellite [15] also shows the exponential relationship [16]. They give a direct measure of photosynthetic potential resulting from the composite property of total leaf chlorophyll, leaf area, canopy cover, and structure. In this case, the NDVI value is closely related to crop yield. It has a direct correlation with LAI, biomass and vegetation cover [17], [18]. Most researches have observed that correlation between NDVI and green biomass yield, therefore NDVI can be used to estimate yield before harvesting [19]. A research was conducted in Indonesia using Landsat, found that the best age of a paddy yield prediction is 63 days after transplanting [16]. Another research in Egypt concludes that the best age of paddy yield prediction is 90 days after transplanting [3]. Another research also in Sri Lanka show the best age of yield prediction is 80 days after transplanting [2].

II. STUDY AREA AND SATELLITE DATA

The paddy area selected for the present study is located in Kiridiwita, Gampaha Sri Lanka. WGS84 coordinates is 79.96983110904694, 7.084400873410644, 79.97372031211853, 7.08574770774949. This paddy area is about 15.92 acres and uniformly transplanting the rice crops in this area. The climatic condition of this study area has major changes in weather occurring during the monsoons from May to August and October to January every year. With the expected heavy rains during these two periods of the year, two seasonal rice crops are always cultivating in this area.

Sentinel-2A satellite used to archive the RS data for this study. Which has Multi-spectral data with 13 bands in the visible, near-infrared, and short wave infrared part of the spectrum. Revisiting every 5 days under the same viewing angles. At high latitudes, Sentinel-2 swath overlap and some regions will be observed twice or more every 5 days, but with different viewing angles [20]. For this study, used five surface reflectance bands. They are 10m spatial resolution surface reflectance band2(490 nm), band3(560 nm), band4(665 nm) and band8(842 nm) and also 20m spatial resolution surface reflectance band11(1610 nm). There are one hundred and seventy-eight images and NDVI values data points from 2015 to 2019 were acquired of AOI through the sentinel-hub cloud(https://www.sentinel-hub.com/). Rice yield data and other required local paddy statistics from 2015 to 2019 were obtained from the Sri Lankan Census and Statistics Department.

III. Methodology

In this study, the main object is forecasting the paddy yield in AOI within the third phase(ripening phase) of paddy stage based on Sentinel-2 RS data. Therefore the first aim is to propose a method to identify the paddy crop stages. For that, image classification approach was used. Since there are very few images, support vector machine classification model was selected [11]. *figure 1* illustrates the approach for paddy stage classification.

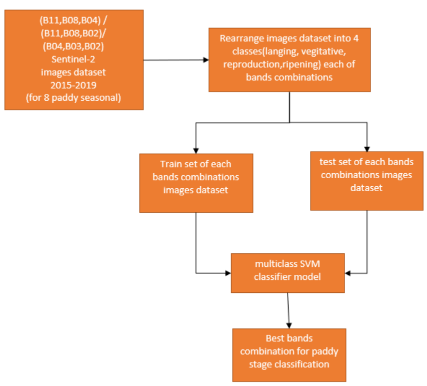

Fig. 1:Paddy Stage Classification Diagram

For building a paddy classification model, surface reflectance band combinations (band4, band3, band2), (band11, band8, band4) and (band11, band8, band2) were selected. (band4, band3, band2) typically known as "True-colour" [20]. which is visible RBG colour. Healthy vegetation reflects NIR(near-infrared) and green bands compare to other wavelengths. And more absorb red and blue bands [21]. SWIR(short wave infrared) used for crop identification [22]. Therefore, (band11, band8, band4) and (band11, band8, band2) were selected rather than used only visible bands. Rearranging image dataset was done according to the Department of Irrigation Yakkala, Sri Lanka water released data. (Figure 2)

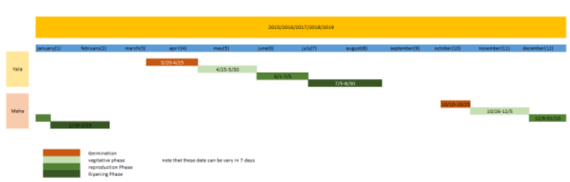

Fig. 2:AOI Paddy Stage Dates

Paddy cultivation in Sri Lanka is based on two monsoons, known as "Yala" and "Maha" seasons. The study was conducted during the two paddy growing seasons and commenced during the same months of each year from 2015 to 2018. The time-series analyses of selected vegetation indices (NDVI) was conducted to build the paddy yield forecasting model.

A. *Vegetation Indices (VI)*

Vegetation Indices (VI) is used to minimize the result of the vegetation canopy radiometric response by factors such as sun-illumination, optical properties of the soil background, view geometries and meteorological factors that were discussed by Leblon (1997). Vegetation Indices (VI) are measures of vegetation "greenness" acquired by combining the results of surface reflectance measurements of the vegetation canopy in different spectral bands [23]. The vegetation indices work based on the principle that vegetation reflects and absorb different amounts of electromagnetic radiation in electromagnetic spectral. Vegetation indices are not characteristic of one particular physical property of the vegetation, but they constitute a combination of properties such as canopy cover, leaf chlorophyll content and leaf area [15]. Many researchers developed a number of different vegetation indexes determine the greenness of plant.

RVI(Ratio Vegetation Index) is the first ratio base vegetation index [24]. It is the ratio between near-infrared and red bands. (equation 1)

$$RVI = \frac{NIR}{RED} \qquad (1)$$

[25] improved the vegetation index and derived Normal Difference Vegetation Index(NDVI) based on near-infrared and red spectral bands as shown in equation 2.

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (2)$$

Green healthy vegetation more absorbs red band and weak absorbs the near-infrared band [21].

B. *Paddy Yield Forecasting Model*

The rice yield forecasting model was derived based on the gradient(coefficient) of NDVI in paddy seasonal. For this experiment 6 paddy seasonals were selected paddy seasons from 2015 to 2018 to build the model and 2 paddy seasonal (2019 Yala & Maha) was selected to do the accuracy assessment. Then the relationship between the gradient(coefficient) of NDVI values and yield was modelled using the recorded average rice yield data from 2015 to 2018. "Greenness" of paddy plants is corresponding to the surface reflectance values of satellite images and it is changing according to the temporal change of paddy plants. Therefore in AOI, obtain the gradient(coefficient) of NDVI values over 90 days time period. This gradient(coefficient) was obtained in linear and polynomial ways.

The best model to represent the relationship between rice yield and NDVI was defined by comparing the results of the accuracy assessment obtained through the linear model. (Figure 3)

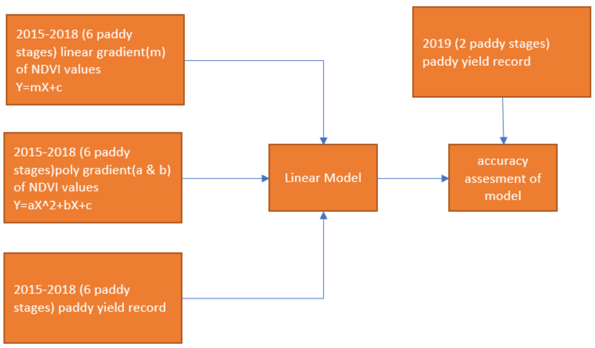

Fig. 3: Yield Forecasting Model

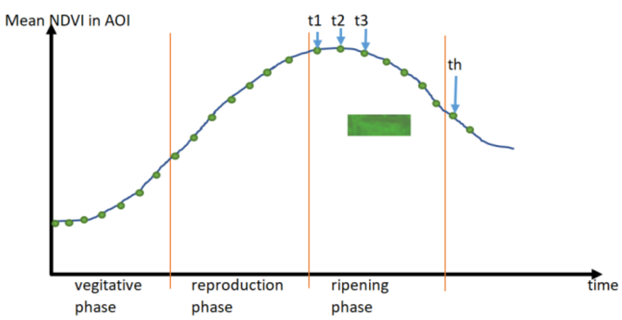

Fig. 3: NDVI Value Change Over The Paddy Seasonal

Figure 3 represents the yield forecasting model and paddy stage classification model works. Once the satellite acquisition image identified the AOI is in the ripening phase then it predicts the rice yield per acre. Find the gradient of linear and coefficient of order two polynomial until the t1 time point and predict the t1time point harvest. Since Sentinel-2A satellite revisiting AOI in every 5 days then the t2 time point came out. Do the same t2 time point and predict the harvest in t2 time point. This goes until the th time point(harvesting time).

IV. Results And Discussions

A. Paddy Stage Classification Model

The cloudy situation eliminates most of the images taken in AOI. Since selected Area is alone with the paddy, it may not worry about identify the paddy crops. Only focus on identify the third phase of paddy crops in AOI. All the downloaded images without cloud cover below 20% in AOI (40% of images) rearrange to 4 classes(Germination, Vegetative phase, Reproductive phase and Ripening phase). Overall there are about 17 images per classes. Since it is very few quantities for analysis of the phase of paddy area, used SVM image classifier. Follow table shows the overall accuracy of each bands combination when classified.

Bands Combination	**F1-Score**
B04, B03, B02	0.77
B11, B08, B04	0.77
B11, B08, B02	0.92

Table 1: F1-Score Of Bands Combinations

According to the above table B11, B08, B02 bands combination shows a more accurate result. So for identifying the third stage of paddy, B11, B08, B02 is selected for optimal bands combination.

B. Rice Yield Forecasting Model

Rice yield forecasting model is based on the gradient of the linear regression and coefficient of a polynomial regression with order 2 in temporal NDVI data with past yield paddy local data.

	Linear Coefficient	**Polynomial Coefficient**	
	X	**X^2**	**X**
Pearson correlation coefficient (ρ)	0.951097	-0.281089	0.698517
Linear regression model	$y = 11.95704437m + 1.29989691$	$y = -0.00148474m + 1.3413055$	$y = 0.00019209m + 1.32071825$

Table 2: Pearson Correlation Coefficient & Models

According to Table 2, the gradient of the linear regression in temporal NDVI paddy seasonal data vs local paddy yield data shows more correlation with $\rho = 0.95$ other than Polynomial coefficients. Therefore $y = 11.95704437m + 1.29989691$ model can be used to predict the paddy yield in AOI.

V. Conclusions And Recommendations

In this current study, initially applied with a very small area about 16 acres alone with paddy crops that have been transplanted the same timeline. The satellite RS data was gathered during 2015 and 2019 for this study. There are six paddy seasonal was selected for this research. Although there were 178 satellite images were download, there were very few images(40%) used for image processing. All other images were covered with cloud and eliminating them. So clouds are really affected to the final result in multi-spectrum satellite-based image processing in agricultural studies. In this study, eliminate all the cloud images, taken in AOI. but for scaling this research cloud effect need to be corrected.

B11, B08, B02 bands combination shows 92% overall accuracy in SVM for paddy stage identification in AOI. In yield forecasting model linear gradient of temporal NDVI seasonal values shows the Pearson correlation coefficient of 0.95 correlation with the past paddy yield data.

ACKNOWLEDGMENT

Sentinel-2 images and NDVI values in AOI were acquired from sentinel-hub cloud GIS platform operated by Sinergise.

REFERENCES

[1] A. a. E. S. Division, "paddy statistics," Department of Census and Statistics, Colombo, Sri Lanka., [Online]. Available: http://www.statistics.gov.lk/agriculture/Paddy%20Statistics/PaddyStats.htm.

[2] P. M. T. S. S. J. S. T. L. Dammalag, "MODIS Satellite Data Based Rice Yield Forecasting Model for Sri Lanka: A Pilot Study in Kurunegala District," *Asian Journal of Geoinformatics.*

[3] M. A. H. A. N.A. Noureldin, "Rice yield forecasting models using satellite imagery in Egypt.," *The Egyptian Journal of Remote Sensing and Space Science,* 2013.

[4] W. T. L. &. F. Kogan, "Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices," *International Journal of Remote Sensing,* 2002.

[5] R.-K. C. Chwen-Min Yan, " Differences in Growth Estimation and Yied Prediction of Rice Crop Using Satellite Data Simulated From Near Ground Hyperspectral Reflectance," *Journal of Photogrammetry and Remote Sensing,* March 2007.

[6] M. I. F. T. B. Sidik Mulyono, " Genetic Algorithm Based New Sequence Principal Component Regression (NS-PCR) For Feature Selection And Yield Prediction Using Hyperspectral Remote Sensing Data," *International Geoscience and Remote Sensing Symposium, 2012* , 2012.

[7] B. P. L. J. H. D. T. G. T. G. F.-Á. I. Ferencz C., " Crop yield estimation by satellite remote sensing," *International Journal of Remote Sensing,* pp. 25, 4113-4149, 2004.

[8] K. B. J.K Shwetank, "Review of rice crop identification and classification using hyper spectral image processing system.," *International Journal of Computer Science & Communication 1,* 2010.

[9] B. A. L. B. A. S. a. R. M. Paul C. Doraiswamy, " Operational Prediction of Crop Yields Using MODIS Data And Products, International Archives of Photogrammetry Remote Sensing and Spatial Information Scince Special Publication," 2007.

[10] J. S. C. L. W.-H. C. Tianxiang Zhang, "Band Selection in Sentinel-2 Satellite for Agriculture Applications," in *Proceedings of the 23rd International Conference on Automation & Computing, University of Huddersfield,Huddersfield, UK, 7-8 September 2017*, China, 2017.

[11] M. I. F. T. B. Sidik Mulyono, "A Paddy Growth Stages Classification Using MODIS Remote Sensing Images with Balanced Branches Support Vector Machines," *ICACSIS 2012.*

[12] P. M. M. M. R. V. Rajitha.K, "Effect of Cirrus Cloud on Normalized Difference Vegetation Index (NDVI) and Aerosol Free Vegetation Index (AFRI): A Study Based on LANDSAT 8 Images".

[13] D. B. D. S. P. Kempeneers, "Wavelet based feature extraction for hyper spectral vegetation monitoring,," *Image and Signal Processing for Remote Sensing* , 2004.

[14] N. H. J Moore, "Examining the development of a potato crop nutrient management trial using reflectance sensing," in *ASAE Annual International Meeting. Las Vegas, Nevada, USA*, Las Vegas, Nevada, USA, 2003.

[15] W. X. Cheng Q., " Mapping paddy rice yield in Zhejiang Province using MODIS spectral index.," 2010.

[16] N. F. H. C. WayanNuarsa I., "Rice Yield Estimation Using Landsat ETM Data and Field Observation," *Journal of Agricultural Science 4* , 2011.

[17] X. B. J. L. D. Z. S. F. S. L. C. W. M. B. Xiao, "Mapping paddy rice agriculture in southern China using multi-temporal MODIS images.," *RemoteSensing of Environment* , pp. 95, 480–492.

[18] R. R. S. Nemani, "Testing a theoretical climate-soil leaf area hydrological equilibrium of forests using satellite data and ecosystem simulation," *Agriculture and Forest Meteorology,* pp. 44, 245–260, 1989.

[19] R. M.S., "Operational yield forecast using AVHRR NDVI data: reduction of environmental and inter-annual variability," *International Journal of Remote Sensing,* pp. 18 (5), 1059–1077, 1997.

[20] w. community, "Sentinel-2," [Online]. Available: https://en.wikipedia.org/wiki/Sentinel-2.

[21] J. C. H Croft, "Leaf Pigment Content," *comprehensive Remote Sensoring,* 2017.

[22] M. M. A. Sruthi Swathandran, "Assessing the role of SWIR band in detecting agricultural crop stress: a case study of Raichur district, Karnataka, India," 2019.

[23] F. G. H. P. J. S. a. A. L. M. Myneni R. B., " The interpretation of spectral vegetation indexes," *IEEE Transactions on Geoscience and Remote Sensing* , pp. 481-486, 1995.

[24] &. M. L. Pearson R.L., "Remote mapping of standing crop biomass for estimation of the productivity of the short-grass," in *8th International Symposium on Remote Sensing of Environment*, 1972.

[25] J. W. H. R. H. S. J. A. D. D. &. H. J. Rouse, " Monitoring the vernal advancement of retrogradation of natural vegetation," Final Report, Type III, NASA/GSFC, Greenbelt, MD, 1974.