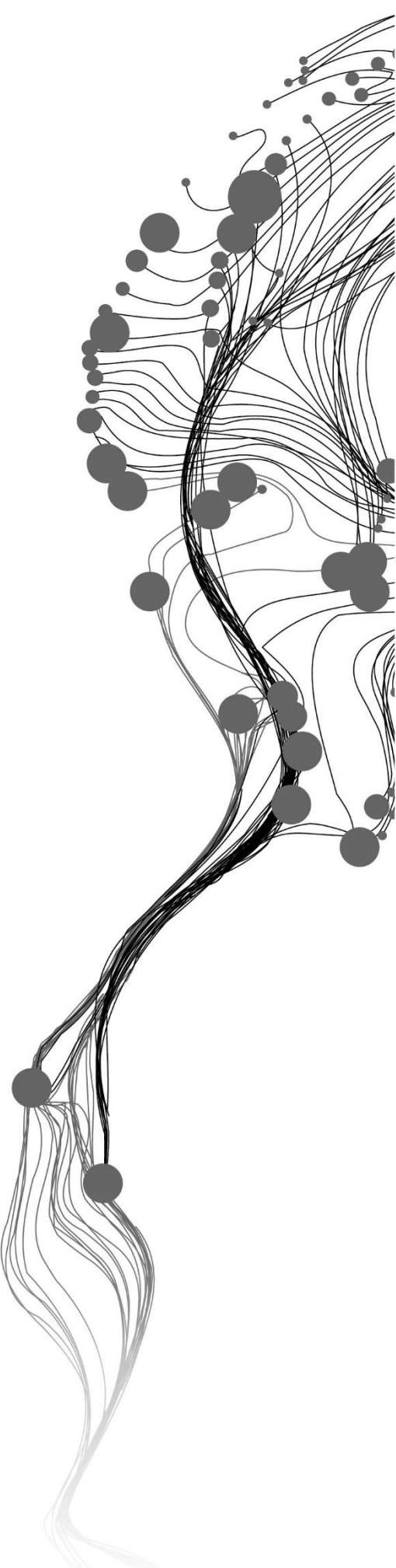


ESTIMATION OF DRY WEIGHT AND HEIGHT OF RICE CROPS USING MACHINE LEARNING ALGORITHMS AND UNMANNED AERIAL VEHICLE (UAV) DATA

FARAH NAFISA ARIADJI
JUNE, 2020

SUPERVISORS:
Prof.dr. Andy Nelson
Dr. Roshanak Darvishzadeh



ESTIMATION OF DRY WEIGHT AND HEIGHT OF RICE CROPS USING MACHINE LEARNING ALGORITHMS AND UNMANNED AERIAL VEHICLE (UAV) DATA

FARAH NAFISA ARIADJI

Enschede, The Netherlands, June, 2020

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.

Specialisation: Natural Resources Management

SUPERVISORS:

Prof.dr. Andy Nelson

Dr. Roshanak Darvishzadeh

THESIS ASSESSMENT BOARD:

dr. Yousif A. Hussin

dr. M Boschetti (Institute for Electromagnetic Sensing of the Environment –
National Research Council of Italy)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Ensuring food security remains one of society's biggest challenges. Rice is a staple food and has an important role in the world's food system. It is engrained in the tradition and culture of many countries. The Philippines is one of the major rice-growing countries and one third of their food consumption is based on rice and its derived products. The high consumption of rice is not balanced with their rice production. This is because there is limited suitable land for rice cultivation, therefore further production growth depends on increasing yield in existing areas. The prediction and estimation of rice yield is necessary to strengthen food security. One way to predict yield is by monitoring and estimating crop parameters, specifically biomass, as they have direct relationship with yield. In addition, crop height as one of the crop parameters is also considered as suitable indicator for plant dry weight estimation and crop growth.

The aim of this study was to accurately estimate the dry weight and height of the rice crop for the early wet season 2016 (2016 EWS) and dry season 2017 (2017 DS) in the Long Term Continuous Cropping Experiment (LTCCE) field in the International Rice Research Institute (IRRI) Experimental Station, Philippines using Unmanned Aerial Vehicle (UAV) data, field data and machine learning algorithms. The UAV dataset consisted of a timeseries of multispectral images of green, red, red edge, near infrared bands and its derived products including point cloud and Digital Surface Model (DSM) data. Field data consisted of 72 samples or subplots within the 1 hectare LTCCE in both the 2016 EWS and 2017 DS. The UAV data was conducted from 23 May to 3 August 2016 and 13 January to 11 April 2017. The machine learning algorithms used in this study were Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF).

In this study, the machine learning algorithms had four different sets of input variables, consisting of the historical total plant dry weight, the average reflectance of four multispectral bands, and vegetation indices. The selection of input variables was based on the correlations between field measured data of plant dry weight and rice height, to average spectral reflectance and vegetation indices.

Low correlations between DSM and point cloud height metrics with field measured crop height were observed which prevented further analysis with the machine learning methods. Reasons for this are discussed and ideas for more representative field measurements of rice crop height are suggested. However, for dry weight (our measure of biomass) we demonstrated that for the 2016 EWS dataset, the SVM method performed best in terms of its accuracy with $R^2 = 0.75$ and RMSE of 639 kg/ha. As for 2017 DS dataset, the best model comes from RF method with $R^2 = 0.88$ and RMSE = 671 kg/ha. We conclude that in this study, SVM and RF algorithm have produced better models compare to ANN algorithms in estimating dry weight in rice with high accuracy by using field measured data and UAV data.

Keywords: rice, UAV, machine learning, crop parameters, sustainable agriculture, food security

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude and praise to the almighty Allah SWT for giving me strength, blessing, and endurance all this time.

I would like to give my sincere gratitude to my supervisors, Prof. Dr. Andy Nelson and Dr. Roshanak Darvishzadeh for their patient guidance, endless supports, and constructive feedbacks and discussions throughout my thesis study. I learned a lot of things and grateful to be under their supervisions. Thank you very much. I would like to thank Dr. Yousif A. Hussin as the Chairman of the thesis assessment board for giving valuable inputs and encouragements during the thesis proposal defense and mid-term presentation.

We thank IRRI for providing the field and UAV data used in this thesis, specifically Mr. Steve Klassen for providing the processed UAV imagery over the LTCCE, documentation, and for clarifications during the thesis research phase. Dr. Roland Buresh, Dr. Sabina Devkota and Dr. Pearl Sanchez for providing data and detailed metadata from their experiments in the LTCCE and clarifications during the thesis research phase. Dr. Pauline Chivenge and Dr. Olivyn Angeles for facilitating access to the data from the LTCCE. The access to and use of these data for this MSc thesis are in line with the Data Access Policy for IRRI's Data on Long-term experiments.

I would like to extend my appreciation to all ITC staffs, Department of Natural Resources (ITC-NRS) and Student Affairs for providing and managing wonderful courses and materials throughout my master study, and thank you for being kind and helpful.

I would like to thank Fanshu Ma for the discussions, ideas and supports, we started from being friends in NRM, until being colleague during our internship and thesis periods. Also thank you for Arsha and Mumtaz, my Indonesian friends for sharing the ups and downs during these two years. My gratitude also goes to the six Ikea girls for their wonderful supports. It is my pleasure to meet and know all of you.

Last and foremost, I would like to thank mama and papa who constantly support, encourage, and pray for me anytime, anywhere. Also for my sister and my brothers who always gave me support and comfort.

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
1.1.	Background.....	1
1.2.	Literature Review.....	3
1.3.	Problem Statement and Justification.....	4
1.4.	Conceptual Framework.....	5
1.5.	Research Identification.....	8
2.	STUDY AREA AND DATA	9
2.1.	Study Area.....	9
2.2.	Field Data.....	10
2.3.	UAV Spectral Data.....	11
2.4.	Software	13
3.	Methodology.....	14
3.1.	Field data and spectral data variables exploration	15
3.2.	Vegetation indices generation	15
3.3.	Statistics of field data and spectral data variables	16
3.4.	Exploratory Analysis of Rice Height Data.....	16
3.5.	Machine Learning Algorithms.....	17
3.6.	Model Evaluation	21
4.	RESULTS	22
4.1.	Preliminary Analyses Results	22
4.2.	Statistics of Field Data and Spectral Data Related to Plant Dry Weight.....	27
4.3.	Machine Learning Models Results.....	28
4.4.	Variable Importance	34
4.5.	Extraction of Digital Surface Model (DSM) for plant height estimation.....	35
4.6.	Derivation of height metrics.....	36
4.7.	Mapping Estimated Dry weight.....	37
5.	Discussions	42
5.1.	Variable selection and variable importance	42
5.2.	Performance of ANN, SVM regression, and RF algorithms in dry weight estimation	42
5.3.	Negative findings of height data	43
5.4.	Limitations	44
5.5.	Recommendation	44
6.	CONCLUSIONS	46

LIST OF FIGURES

Figure 1. Average daily consumption per capita by several food groups in Philippines, 2008.....	2
Figure 2. The growth phase of transplanted rice	3
Figure 3. Conceptual diagram of the study	5
Figure 4. The study area is located in LTCCE at IRRI, Los Baños, Philippines.	9
Figure 5. The layout field of LTCCE in year 2016 EWS and 2017 DS.....	10
Figure 6. The layout of field dry weight measurement within each subplot	11
Figure 7. The wavelength response of each sensor in multiSpec 4C camera	12
Figure 8. Flowchart of the study.....	14
Figure 9. Anatomy of artificial neuron.....	17
Figure 10. An example of the architecture of ANN-MLPs model	18
Figure 11. Boxplots showing the variation of total plant dry weight of rice at different dates.	23
Figure 12. Boxplot of rice crop height of the three different varieties in 2016 EWS (a) and 2017 DS (b)..	23
Figure 13. Correlation between plant height and total plant dry weight in the 2016 EWS and 2017 DS	24
Figure 14. Mean reflectance of rice fields of green, red, red edge, and near infra-red bands throughout the 2016 EWS.....	25
Figure 15. Mean reflectance of rice fields in green, red, red edge, and near infra-red bands throughout the 2017 DS	26
Figure 16. Correlations between total plant dry weight and multispectral bands in 2016 EWS (a) and 2017 DS (b)	27
Figure 17. The correlations of total plant dry weight and vegetation indices in 2016 EWS (a) and 2017 DS (b).....	28
Figure 18. Model 1 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 data with their R ² and RMSE. Red dotted line is the 1:1 line.....	29
Figure 19. Model 2 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 data with their R ² and RMSE. Red dotted line is the 1:1 line.....	30
Figure 20. Model 3 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 EWS data with their R ² and RMSE. Red dotted line is the 1:1 line.	30
Figure 21. Model 4 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 EWS data with their R ² and RMSE. Red dotted line is the 1:1 line.	30
Figure 22. Model 1 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R ² and RMSE. Red dotted line is the 1:1 line.....	32
Figure 23. Model 2 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R ² and RMSE. Red dotted line is the 1:1 line.....	32
Figure 24. Model 3 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R ² and RMSE. Red dotted line is the 1:1 line.....	33
Figure 25. Model 4 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R ² and RMSE. Red dotted line is the 1:1 line.....	33
Figure 26. Field measured height vs height derived using DSM data (average value).	35
Figure 27. The maps of field measured dry weight and estimated dry weight using SVM algorithm in 2016	38
Figure 28. The difference map and the percentage difference map.....	39
Figure 29. The maps of field measured dry weight and estimated dry weight using RF algorithm in 2017.40	
Figure 30. The difference map of dry weight in 2017 and the percentage difference map	41

LIST OF TABLES

Table 1. Related studies estimating crop parameters	6
Table 2. Dates of dry weight field measurements in 2016 EWS and 2017 DS	11
Table 3. The details of UAV image acquisitions and the image resolutions.....	12
Table 4. The details of DSM and point cloud data	13
Table 5. The list of software used in this study	13
Table 6. Vegetation indices used in this study.....	16
Table 7. The list of independent variables as input for the model 1.....	19
Table 8. The list of independent variables as input into model 2.....	20
Table 9. The list of independent variables as input to model 3	20
Table 10. The list of independent variables as input to model 4	21
Table 11. The summary of the calibration results for four models with 2016 EWS dataset	29
Table 12. The summary results of ML algorithms when calibration and validation datasets of 2016 EWS were used.....	31
Table 13. The summary of the calibration results for four models with the 2017 DS dataset.	31
Table 14. The summary results of ML algorithms when calibration and validation datasets of 2017 DS were used.....	34
Table 15. The list of variable importance of 2016 EWS data.....	34
Table 16. Height metrics derived from point cloud data	36
Table 17. Correlations between point cloud metrics and height obtained from field measurements	37

1. INTRODUCTION

1.1. Background

Two of the challenges the world is facing today are growing population and consumption, which are increasing the global demand for resources (Foley et al., 2011; Godfray et al., 2010). These phenomena are resulting in societal and environmental problems, such as hunger and poverty. According to the Food and Agriculture Organization of the United Nations, there are 820 million people in the world who suffer from hunger and undernourishment (FAO, IFAD, UNICEF, WFP, & WHO, 2018). Hunger, undernourishment, and poverty are three global challenges that relate to food security. Based on the World Food Conference in 1996 (FAO, 1996) the definition of food security is: “Food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life”. Ensuring food security remains one of society’s biggest challenges.

Rice has an important role in the world’s food system. More than a third of the global population, mainly in Asia, consumes rice as their staple food (Barker, Herdt, & Rose, 1985). According to Papademetriou, Dent, and Herath (2000), the Asia-Pacific region produces and consumes almost 90% of the world’s rice. Rice consumption also has increased in sub-Saharan Africa, the Caribbean, and in Latin America in recent decades. Particularly in low and lower-middle-income countries, rice is one of the most important foods, covering 19% of the total crop area harvested (GRiSP, 2013). Rice is ingrained in the tradition and culture of many countries. The 2008 rice price crisis is one example period where rice-producing countries limited their exports, which resulted in rice scarcity and a price surge which affected all rice consumers, especially the poor. The crisis partially triggered rice-producing and rice-consuming countries to become more self-sufficient and achieve food security (FAO, 2010; GRiSP, 2013). Therefore, rice plays an important role in food security and is considered as a political commodity, with important social and economic aspects to its production, trade and consumption (FAO, 2006).

The Philippines is one of the major rice-growing countries where rice has been cultivated for at least 5,000 years. Rice is the main staple food, and the demand for rice continues to increase with the rapidly growing population (Moya, Dawe, & Casiwan, 2006). By the year 2012, Philippines was the eighth biggest rice producer in the world with 18 million tons of rice produced (United Nations Development Programme (UNDP), 2015). Figure 1 shows Filipino food consumption patterns with rice accounting for over one third of rice production.

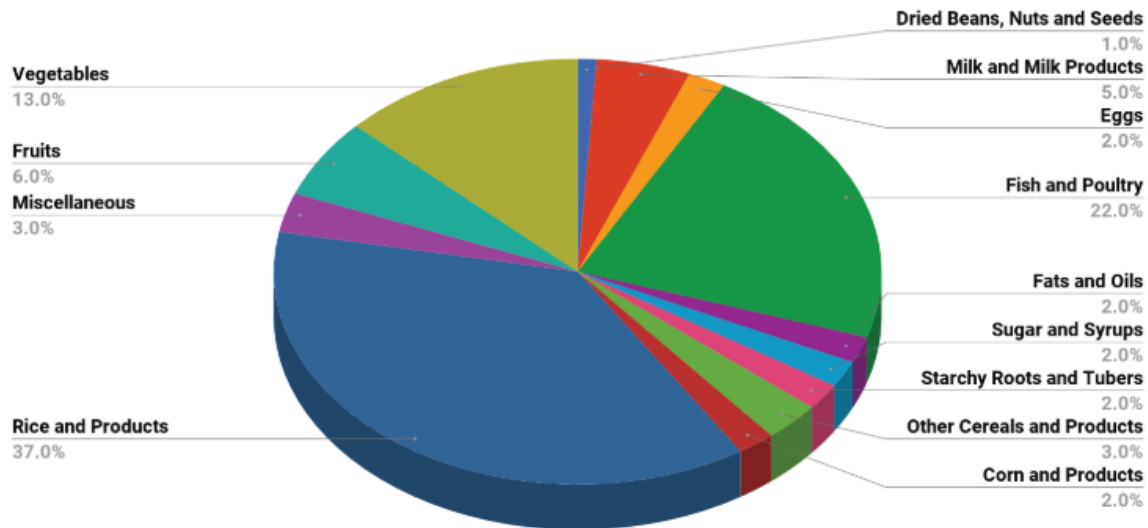


Figure 1. Average daily consumption per capita by several food groups in Philippines, 2008. Source: (Ponce & Inocencio, 2017)

However, the land that is suitable for cultivating rice in the Philippines is limited to a few lowland areas, amounting to only 0.13 ha per capita (OECD, 2017). With little opportunity to expand the area, further production growth depends on producing higher yields in existing production areas. Although yield has increased from 2.8t/ha in 1995 to 3.6t/ha in 2010 and 4.0t/ha in 2017 (FAO, 2019), it is still substantially below the potential yield from modern rice varieties and below that from neighbouring countries such as Vietnam and Indonesia (GRiSP, 2013). There are other constraints that limit rice production in the Philippines. Extreme weather such as typhoons that often occur lead to heavy rainfall and drought often occurs in rainfed areas (GRiSP, 2013).

Prediction and estimation of rice yield are necessary for strengthening food security. Having better yield prediction can help in decision making for optimal planting schedules, marketing, transporting, and storage (Tennakoon, Murty, & Eiumnoh, 1992). One way to predict yield is by monitoring and estimating crop parameters, specifically dry weight, as it has a direct relationship with yield (Jin & Liu, 1997). Moreover, Lati, Filin, and Eizenberg (2013) demonstrated in their study that dry weight values, height, and leaf cover area (LCA) are able to model the plant's growth. Scully and Wallace (1990) also showed that dry weight and dry weight growth are able to indicate potential crop yield. In addition, another crop parameter, crop height, is considered as a suitable indicator for dry weight estimation and crop growth because crop height significantly affects yield potential (Ehlert, Adamek, & Horn, 2009; Lati et al., 2013; Ndikumana et al., 2018). Radloff and Mucina (2007) estimate dry weight based on the regression relationship between height measurements and dry weight. Figure 2 shows the growth duration of transplanted rice.

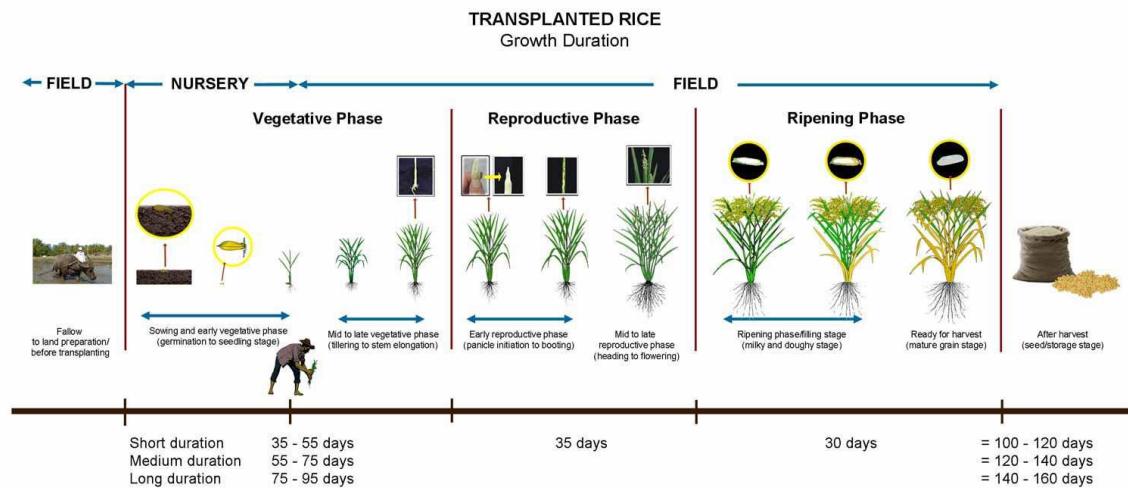


Figure 2. The growth phase of transplanted rice (source: <http://www.knowledgebank.irri.org/step-by-step-production/pre-planting/crop-calendar>)

1.2. Literature Review

Field measurement of plant dry weight and height are time consuming and costly since they need to cover a representative sample of rice crops and environments, and require repeated measures through the season to capture changes through the growth cycle. Remote sensing data is an alternative source of information to estimate crop dry weight and height. Various passive and active remote sensing sensors on different platforms; ground, unmanned aerial vehicle (UAV), airborne, and satellite, have been utilized for monitoring these two crop parameters (J. V. Bendig, 2015; Luo et al., 2015; Ndikumana et al., 2018; Reisi-Gahrouei, Homayouni, McNairn, Hosseini, & Safari, 2019; Tilly et al., 2014; Xie, Sha, Yu, Bai, & Zhang, 2009). A review of the literature is summarised in Table 1 and several of the most relevant studies for this research are described in more detail in this section.

UAVs have become a promising remote sensing platform to capture detailed imagery for agricultural fields and for estimating crop parameters (Malambo et al., 2018). The use of UAVs in farming for optimizing management has become a major opportunity to obtain field data in a simple and fast way (Bansod, Singh, Thakur, & Singhal, 2017). UAV data usually has a very high spatial resolution, ranging from 0.5 – 10 cm with a field of view of 50 – 500 m (Candiago, Remondino, De Giglio, Dubbini, & Gattelli, 2015).

UAV data has been used in many studies for crop monitoring. Vega, Ramírez, Saiz, and Rosúa (2015) used UAV data to monitor sunflower properties like leaf area index, water stress, nitrogen content, the yield of grains, and dry weight during the growing season. Chang et al. (2017) generated ortho-mosaic and 3D point cloud information using a Structure from Motion (SfM) algorithm on 200 overlapping images from an altitude of 50 meters, and obtained a very low root mean square error (RMSE) between the estimated height and field measured height. Schirrmann et al. (2016) monitored biophysical parameters of winter wheat, including height, dry weight, LAI, as well as nitrogen status, using UAV data in an 11 ha field and observed high correlation between biophysical parameters and image variables (such as cover coverage, band ratios, and plant height), $R^2 = 0.70 - 0.97$. Li et al. (2016) showed that maize height derived from UAV images was comparable with the field measurements with an RMSE of 0.11 m. Several studies have successfully conducted similar measurements with different crop types, including barley (J. Bendig et al., 2014), soybean (Yu et al., 2016), or winter wheat (Schirrmann et al., 2016). Estimation of crop height is essential in precision

agriculture practices since it can be used to predict yield, irrigation schedule, and application of pesticide and fertilizer (A. Chang et al., 2017; Lati et al., 2013).

Using spectral reflectance extracted from remote sensing data, vegetation indices (VIs) have been used to qualitatively and quantitatively evaluate crop parameters like dry weight and crop height (Bannari, Morin, Bonn, & Huete, 1995; Jackson & Huete, 1991). Some studies have proved that VIs are able to estimate dry weight (Cho, Skidmore, Corsi, van Wieren, & Sobhan, 2007; Devia et al., 2019; Stavrakoudis, Katsantonis, Kadoglidou, Kalaitzidis, & Gitas, 2019). However, there are some limitations of using VIs, such as VIs have not been calibrated for all crop species, meaning the same coefficients are applied for different crops (Gholizadeh, Robeson, & Rahman, 2015).

Recently, machine learning techniques (Hastie, Tibshirani, & Friedman, 2008) - that can help in finding rules and patterns in complex multivariate data (McQueen, Garner, Nevill-Manning, & Witten, 1995) - have been applied to find relationships between VI information and field measurements of crop parameters. Moeckel et al., (2018) applied Random Forest (RF) and Support Vector Machine (SVM) methods to predict the height of crops using UAV imagery during an entire growing period resulting in a pseudo- R^2 of 0.87 (RMSE = 5.91), 0.91 (RMSE = 7.36), and 0.89 (RMSE = 2.31) for tomato, eggplant, and cabbage, respectively for SVM and a pseudo- R^2 of 0.89 (RMSE=5.49), 0.93 (RMSE = 6.86) for RF. Ndikumana et al., (2018) also performed similar research, estimating rice height and dry weight using Sentinel-1 which resulted in a high correlation between estimated measurement and in-situ measurement from dual-polarization information. Han et al. (2019) also modelled maize dry weight based on Multiple Linear Regression (MLR), Artificial Neural Network (ANN), SVM, and RF algorithms using UAV data. They found that RF gave the most balanced results with a RMSE = 0.495 and $R^2 = 0.944$ for their training set and a RMSE = 1.2 and $R^2 = 0.699$ for their test set. A study by Yue, Feng, Yang, and Li (2018) utilised MLR, ANN, RF, SVM, Decision-tree regression (DT), Boosted binary regression tree (BBRT), partial least squares regression (PLSR), and principal component regression (PCR) to estimate winter wheat dry weight using near spectroscopy hyperspectral data and obtained accurate results ($R^2 = 0.89$ and RMSE = 1.20 t/ha). Ali, Cawkwell, Dwyer, and Green (2017) developed a model for dry weight estimation in grassland based on MLR, ANN, and adaptive neuro-fuzzy interference system (ANFIS) algorithms and Moderate Resolution Imaging Spectroradiometer (MODIS) Terra data. Reisi-Gahrouei et al. (2019) estimated dry weight using UAVSAR data and comparing multilinear regression and ANN algorithms in some crops, resulting in ANN has higher accuracy with R^2 ranging from = 0.72 to 0.92, RMSE = 13.48 g/m² (soybean), 56.55 g/m² (canola), 196.71 g/m² (corn). A study by Wang et al. (2019) on growth stage development included height as a parameter when comparing ANN, SVM, RF, and partial least squares regression (PLSR) and showed that RF had the highest accuracy with RMSE of 20.21%. Devia et al. (2019) estimated dry weight in rice crops in both dry and wet seasons, using UAV with multispectral sensor, particularly near infra-red (near-infrared). The study located in the International Centre for Tropical Agriculture (CIAT) experimental station in Colombia and they considered three different crop stages. They applied multivariable regressions to model seven VIs and the relations to dry weight and compare the regression method to ANN method. The results showed in vegetative and reproductive stages, ANN method performed better with $R^2 > 0.8$, but the performance decrease in ripening stage. Overall, multivariable regressions are more reliable than ANN, as ANN needs sufficient training data to be utilized.

1.3. Problem Statement and Justification

As the result of the literature review of twenty studies (see Table 1), six studies estimated dry weight and/or height of rice crop, and three of them were using UAV images as source of input data, but only one study used a machine learning algorithm.

Based on the aforementioned studies, estimations of dry weight and height in rice crop through the growing season have not yet been studied using UAV images with different machine learning approaches. In this study, we will utilize field data and UAV data from four different sensors (red-green-blue (RGB), four-band multispectral (R-G-RedEdge-Near-infrared), RGNIR, and thermal) and three different machine learning algorithms (Artificial Neural Network, Support Vector Machine, and Random Forest) to estimate dry weight and height of the rice crop through two consecutive growing seasons. The study area is in the International Rice Research Institute (IRRI) Experimental Station, Los Baños, Philippines and all data were collected and provided by IRRI.

1.4. Conceptual Framework

Figure 3 shows the conceptual diagram of this study. The geographical boundary of the system is Los Baños, Philippines, where the IRRI Experimental Station is located. The elements of the system consist of the rice fields in IRRI Experimental Station, IRRI, the villages, and also the irrigation system. Key characteristics of the rice fields include growing seasons, soil condition, fertilizer water cycle, and rice phenology that are related to rice crop growth. As sub-element, there are the two crop parameters of dry weight and height that will be observed using UAV images and measured with field measurements and a machine learning approach will find relations between these two. The UAV-derived data consists of the spectral reflectance of multispectral bands, point cloud, vegetation indices, and DSM. The results of the study are expected to be immediately useful for IRRI scientists who are using UAV imagery to efficiently collect data on field experiments to understand crop growth under different management conditions. The eventual benefits of their research are directed to producers and consumers in terms of higher and more efficient rice productivity.

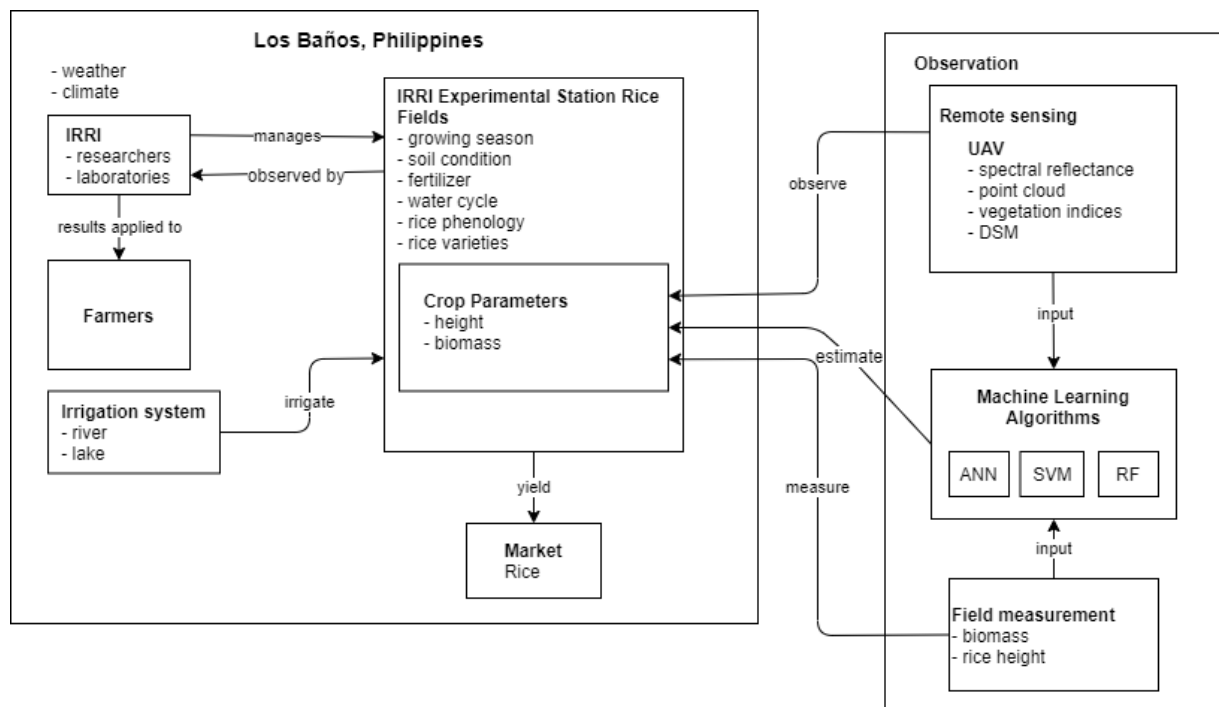


Figure 3. Conceptual diagram of the study

Table 1. Related studies estimating crop parameters

NO	CROP	FOCUS	TYPE OF OBSERVATION	METHOD	RESULTS	REFERENCE
1	Rice	Dry biomass and height	Sentinel 1	multiple linear regression (MLR), SVR, RF	<ul style="list-style-type: none"> - most accurate algorithm to estimate height: RF $\rightarrow R^2 = 0.92$, RMSE = 16% (7.9 cm) - using RF, biomass estimation has high accuracy with $R^2 = 0.9$, RMSE = 18% (162 g/m²) 	Ndikumana et al., 2018
2	Rice	Above ground biomass	Spectrometer	Soil adjusted vegetation indices (Vis), optimized narrow band RVI and NDVI, Optimum Multiple Narrow Band Reflectance (OMNBR) models	<ul style="list-style-type: none"> - the most accurate method is six-band OMNBR models ($R^2 = 0.44-0.73$) explained 21–35% more AGB variability 	Gnyp et al., 2014
3	Rice	Biomass	UAV-RGB	<ul style="list-style-type: none"> - simple linear regression (SLR), simple exponential regression (SER), RF (for testing) - Triangulated Irregular Network (TIN) from point cloud 	<ul style="list-style-type: none"> - using RF method, the combination of multispectral, structural, and metrological features for rice AGB estimation improved the result with $R^2 = 0.86$, RMSE = 178.37 g/m² - For the other TIN-based structural features, SLR has higher R^2 than SER 	Jiang et al., 2019
4	Rice	LAI (effect on growth stage development)	Spectroradiometer - Hyperspectral	PLSR, SVR, RF, ANN	<ul style="list-style-type: none"> - RMSE = 20.21, $R^2 = 0.39$ - RF and SVR showed better performance than PLSR and ANN 	Wang et al., 2019
5	Rice	Above ground biomass (AGB)	UAV-Multispectral NIR	multivariable regression of 7 vegetation indices compare to ANN	<ul style="list-style-type: none"> - In the vegetative and reproductive stages, ANN performed better with $R^2 > 80\%$ - overall, multivariable regression is more reliable than ANN - the proposed method has $R^2 = 0.76$ in biomass estimation 	Devia et al., 2019
6	Rice	Plant height, AGB, N concentration, N uptake, grain yield, and harvest index	UAV-Multispectral	MLR, vegetation indices	<ul style="list-style-type: none"> - At the <u>tillering</u> stage, high accuracies ($R^2 \geq 0.8$) were achieved for N uptake and biomass. - Similar result for estimation of yield, N concentration, N uptake, biomass, and plant height 	Stavarakoudis et al., 2019
7	Winter wheat	N content, NNI, height, biophysical variables (LAI, fresh and dry biomass)	UAV-RGB	linear regression (PCA), image variables (RGB ratios, plant height, crop coverage)	<ul style="list-style-type: none"> - Biophysical variables and image variables: $R^2 = 0.77-0.97$. - N model has $R^2 = 0.65$ - NNI RMSE = 0.10-0.11 	Schirrmann et al., 2016
8	Winter wheat	Above ground biomass	Spectrophotometer	54 VIs and 8 statistical regression techniques (5 machine learning and 3 conventional regression)	<ul style="list-style-type: none"> - PLSR and MLR perform best in stability and most suitable with high accuracy for few samples - RF is very robust against noise and most suitable for RS repeated data - PLSR has the highest accuracy ($R^2 = 0.89$, RMSE = 1.20t/ha) 	Yue, Feng, Yang, & Li, 2018

9	Wheat, oat	biomass parameters (canopy height, canopy water content, dry matter, volumetric wetness of the land)	High-dimensional (multi-day, dual polarization, active, passive) remote sensing data	ANN	- Results (in form of plots) show ANN is capable in inversion retrievals of parameters to train active and passive data	Jin & Liu, 1997
10	Maize	Biomass and canopy height	UAV-RGB	- VIs variables - SfM, point clouds - stepwise linear, RF	- H_{canopy} and AGB were predicted by stepwise linear (SWL) regression model ($R^2 = 0.88$, RMSE = 6.40%) and RF ($R^2 = 0.78$, RMSE = 16.66%) - between UAV and field measurement: mean error (ME) = 0.11 m for H_{canopy} , and 0.05 kg/m ² for AGB	Li et al., 2016
11	Maize	Above ground biomass	UAV-RGB and UAV-Multispectral	MLR, SVM, ANN, and RF	- RF produced the highest accuracy → RMSE = 0.495, $R^2 = 0.944$ for training set; RMSE = 1.2, $R^2 = 0.699$ for test set	Han et al., 2019
12	Maize, sorghum	Height	UAS and TLS	Structure for Motion (SfM), linear regression	- Measurement between UAS& TLS → $R^2 = 0.88-0.97$, RMSE = 0.01–0.02 m - Ground surface and canopy comparison → $R^2 = 0.60-0.77$, RMSE = 0.12–0.16 m	Malambo et al., 2018
13	Sorghum	Height	UAS (UAV-RGB)	Crop Height Model (CHM) (derived from point clouds)	- between field measurements and the model → RMSE = 0.33 m	Chang, Jung, Maeda, & Landivar, 2017
14	Grass	Biomass	MODIS Terra	MLR, ANN, adaptive neuro-fuzzy inference system (ANFIS)	- ANFIS has highest accuracy compare to MLR and ANN → $R^2_{Moorepark} = 0.85$, RMSE _{Moorepark} = 11.07; $R^2_{Grange} = 0.76$, RMSE _{Grange} = 15.35	Ali, Cawkwell, Dwyer, & Green, 2017
15	Grass	Biomass	Field measurement	The Multipoint Minidisk Meter method	- between biomass and height values → $R^2 = 0.90 - 0.96$	Radloff & Mucina, 2007
16	Barley	Fresh and dry biomass	UAV-RGB	Crop Surface Model (CSM), regression	- Estimated and field measurement heights → $R^2 = 0.92$ - Height estimation and biomass → $R^2 = 0.81$ (fresh) and $R^2 = 0.82$ (dry)	Bendig et al., 2014
17	Sugar beet	Modelling crop surface and volume	TLS	CSM and CVM	- RMSE = 0.019	Hoffmeister et al., 2009
18	Eggplant, tomato, cabbage	Height	UAV-RGB	SfM, RFR, SVR, PLSR	- RF has the highest accuracy compare to SVM and PLSR - Height estimation → R^2 for all crops ranging from = 0.89 – 0.97, RMSE ranging from = 1.3 – 6.89	Moeckel et al., 2018
19	Corn, sunflower, cotton	plant height, leaf cover area (LCA), biomass	Close range camera	3D stereovision model, regression	- Height RMSE = 4%, LCA RMSE = 4.5% - Measured biomass and estimated volume → $R^2 = 0.94$, RMSE = 4%	Lati, Filin, & Eizenberg, 2013
20	Canola, corn, soybean	Biomass	Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR)	MLR, ANN	- ANN has higher accuracy than MLR - The R^2 ranging from = 0.72 – 0.92, RMSE = 13.48 g/m ² (soybean), 56.55 g/m ² (canola), 196.71 g/m ² (corn) - All crops pooled → $R^2 = 0.89$, RMSE = 135.31 g/m ²	Reisi-Gahrouei et al., 2019

1.5. Research Identification

The aim of this study is to accurately estimate the dry weight and height of the rice crop using UAV data field measurements and machine learning methods based on observations and measurements obtained during the wet season 2016 (2016 EWS) and dry season 2017 (2017 DS) in the IRRI Experimental Station, Philippines.

1.5.1. Research Objectives

In order to achieve the aim, the objectives of the study are as follow:

1. To evaluate different machine learning algorithms - ANN, SVM, and RF - for the estimation of dry weight in the rice crop in wet and dry seasons using UAV data.
2. To evaluate different machine learning algorithms - ANN, SVM, and RF - for the estimation of rice crop height in wet and dry seasons using UAV data.
3. To map the dry weight and height of the rice crop in wet and dry seasons using the most accurate algorithm (in terms of R^2 and RMSE).

1.5.2. Research Questions and Hypothesis

Several research questions were formulated to address the objectives:

1. Which machine learning algorithm has the best accuracy (in terms of R^2 and RMSE) for estimation of dry weight in the wet and dry season?

Hypothesis: Random Forest is the most accurate algorithm for estimating dry weight in both wet and dry seasons.

2. Which machine learning algorithm has the best accuracy (in terms of R^2 and RMSE) for estimation of rice height in the wet and dry season?

Hypothesis: Support vector machine is the most accurate algorithm for estimating rice height in both wet and dry season.

2. STUDY AREA AND DATA

2.1. Study Area

The study takes place in the Long-Term Continuous Cropping Experiment (LTCCE) site maintained by IRRI, located in Los Baños, Philippines. The LTCCE was established in 1962 with the aim to have a high annual rice yield from a unit area of land by testing various rice varieties, irrigation, proper timing and use of fertilizer, and agricultural practices (Santiaguel, 2014). The LTCCE has been continuously cultivated with two or three crops per year since 1962 providing information on yield trends in intensively cultivated rice cropping systems. The LTCCE covers an area of one-hectare with a centre coordinate of $14^{\circ}10' 05.7''\text{N}$ and $121^{\circ}15' 21.2''\text{E}$ (Figure 4). There are two sets of data from this site that this study will focus on, one from the year 2016 for the early wet season (2016 EWS) and the other from the year 2017 for the dry season (2017 DS).

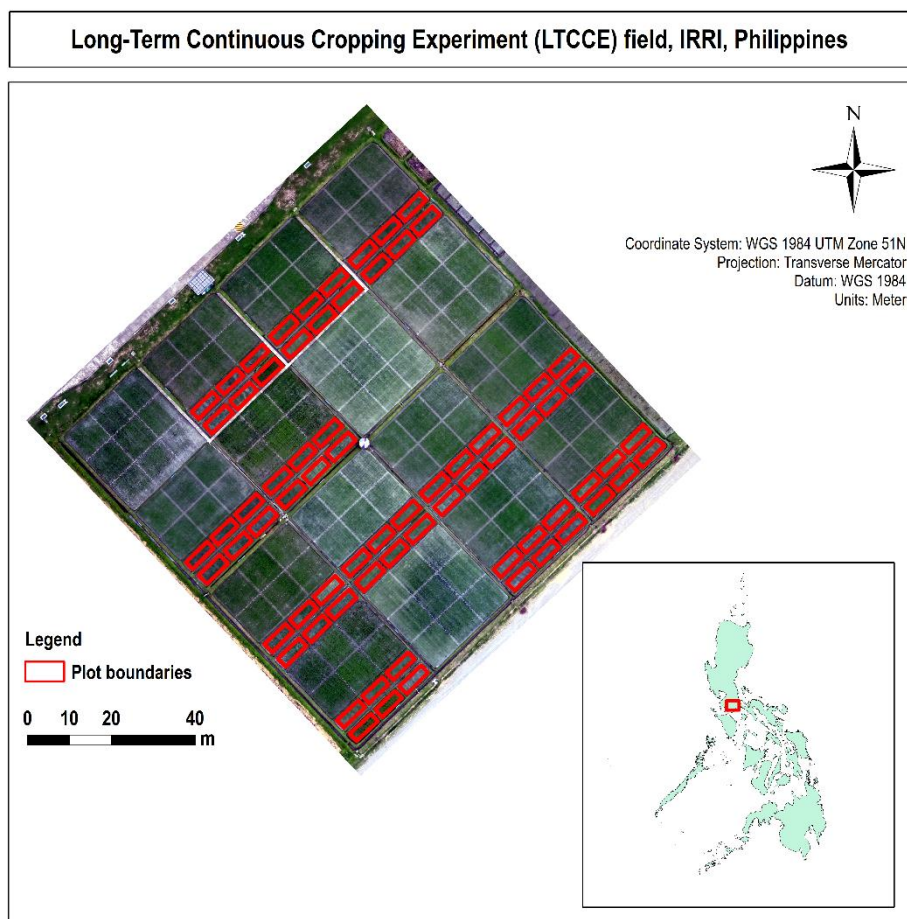


Figure 4. The study area is located in LTCCE at IRRI, Los Baños, Philippines. Red polygons show the subplots studied in this research. The image used is an RGB image taken in 23 March 2016.

The LTCCE site is divided into 108 main plots of $8\text{m} \times 8\text{m}$ which are used to measure yield and 72 subplots of $4\text{m} \times 8\text{m}$ used to measure vegetation biophysical parameters (these plots are shown in white in Figure 5). The LTCCE contains different combinations of rice varieties and nitrogen applications. Based on Figure 5, this study will focus on 72 white subplots where biophysical parameters were measured, such as plant dry weight and rice height for three rice varieties and six nitrogen applications. The details of the field data measurements will be explained in the next section.

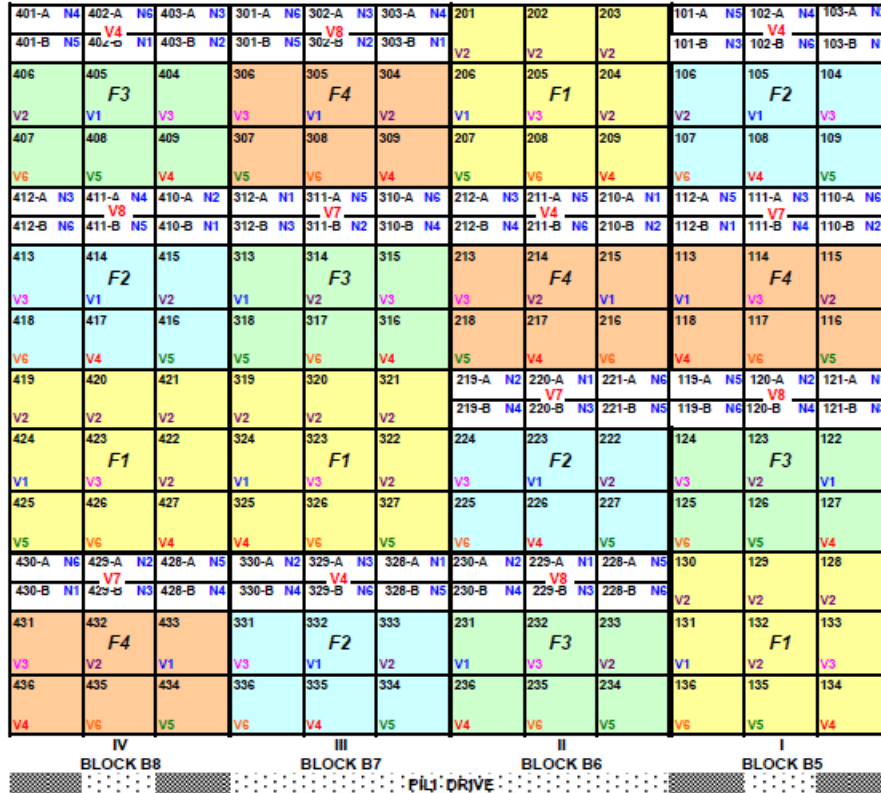


Figure 5. The layout field of LTCCE in year 2016 EWS and 2017 DS (both seasons have the same field layout) (IRRI, 2019). F1 – F4 are the main plots with different rates of nitrogen; v1 – v8 are the rice varieties; The N1 – N6 show the nitrogen application rate for each subplot.

2.2. Field Data

The variables used in this study are the dry weight and height of rice crop. This section will describe the measurement of both variables by IRRI.

Field data collection was conducted four times during the season for plant dry weight measurements and once for rice height. Dry weight data includes dry dry weight of dead-leaf, green leaf, stem, panicle, and total plant dry dry weight. These measurements were available for all subplots. Table 2 shows the dates for operation and sampling of plant dry weight measurements in each season based on days after transplanting (DAT). As for rice height, it was measured once in each season and the measurement took place between 4 August 2016 – 11 August 2016 and 13 April 2017 – 18 April 2017.

Table 2. Dates of dry weight field measurements in 2016 EWS and 2017 DS

2016 EWS		2017 DS	
Calendar date	Days after transplanting (DAT)	Calendar date	Days after transplanting (DAT)
26 May 2016	22	27 Jan 2017	23
15 June 2016	42	16 February 2017	43
28 June 2016	55	2 March 2017	57
5 July 2016	62	13 March 2017	68

Destructive sampling was conducted in each subplot of $3.4\text{m} \times 7.6\text{m}$ (smaller than the subplot to exclude edge effects from plants at the border of the plot), which contained 17×38 rows of rice hills at a spacing of $20\text{cm} \times 20\text{cm}$. Plant dry weight was measured separately for green leaf, dead leaf, and stem. Panicles were sampled at the period between booting and heading stage. The plant samples were oven-dried for three days at a temperature of 80°C . However, the calculation involving plant dry weight in this study used the accumulation of green leaf, dead leaf, and stem for the first three sampling measurement, plus panicle on the last day of the destructive sampling, which is termed “total plant dry weight” in further analysis. Figure 6 shows the location of the destructive sampling or field measurement of the dry weight in one subplot.

As for the rice crop height, it was measured at four locations within each subplot by stretching up the flag leaf (the tallest leaf) of the rice plant. The measurements were obtained at the harvest date, and the mean of the four measurements was reported as the average height of the subplot. There were no coordinates available for the locations of these measurements, so subplot averages were used for further analysis

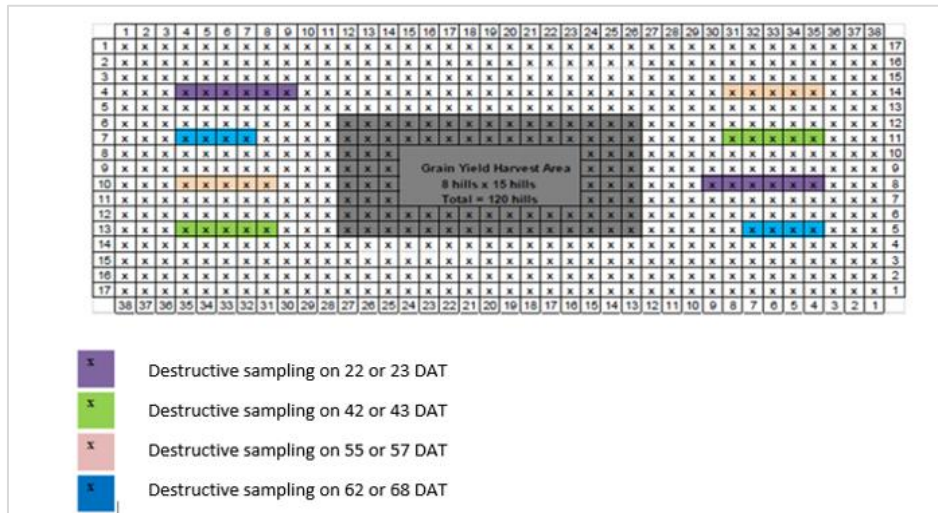


Figure 6. The layout of field dry weight measurement within each subplot (IRRI, 2019)

2.3. UAV Spectral Data

The UAV multispectral data acquired in two seasons, the wet season in 2016 and the dry season in 2017, are the primary remote sensing data for this study. The data were collected using a fixed-wing UAV – Sensefly multiSpec4C. The multispectral image consists of red-green-red edge-near infrared bands have already been processed into mosaics using Pix4D software and spatially corrected with the GCPs. Figure 7

illustrates the wavelength response of multiSpec 4C camera. More details about the UAV multispectral data are shown in Table 3 for the specific dates of acquisitions and pixel resolutions.

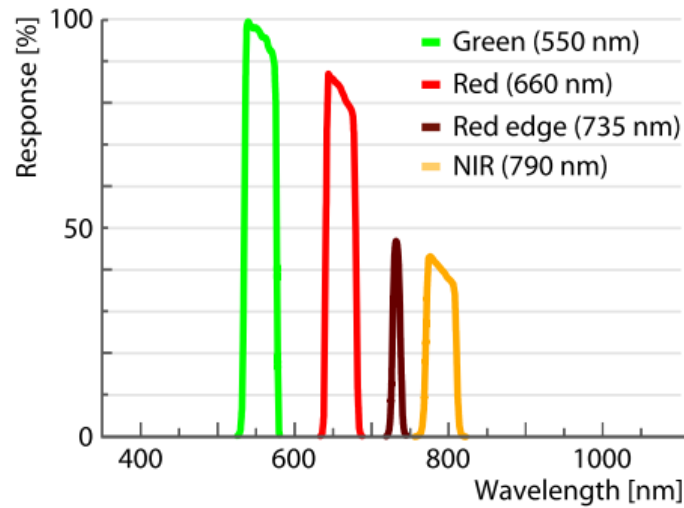


Figure 7. The wavelength response of each sensor in multiSpec 4C camera (senseFly, 2014)

Table 3. The details of UAV image acquisitions and the image resolutions

Date of acquisition	Pixel resolution (cm)	Date of acquisition	Pixel resolution (cm)
2016-05-23	7.6	2017-01-13	7.28
2016-05-25	7.4	2017-01-19	7.05
2016-06-01	7.4	2017-02-08	7.49
2016-06-08	7.6	2017-02-16	7.34
2016-06-14	7.1	2017-02-23	7.37
2016-06-21	7.3	2017-03-02	7.19
2016-06-28	7.3	2017-03-08	7.6
2016-07-05	7.7	2017-03-15	7.6
2016-07-06	7.4	2017-03-22	6.2
2016-07-22	7.2	2017-03-29	6.78
2016-07-27	7.1	2017-04-05	6.98
2016-08-03	7.4	2017-04-11	7.1

Due to the field measured height only took one measurement, the UAV-derived height data also had to choose from one dataset. Table 4 shows the details of the DSM and point cloud data that were used in this study. The 2016 EWS took the DSM and point cloud data from DAT 62, and the 2017 DS took the DSM and point cloud data from DAT 68.

Table 4. The details of DSM and point cloud data

Season	Date	Resolution of DSM (cm/pixel)	Number of densified points	Average density (per m ³)
2016 EWS	5 July 2016	2.42	1980855	190.55
2017 DS	15 March 2017	2.46	1505480	217.91

2.4. Software

The software used in this study are listed in Table 5 below.

Table 5. The list of software used in this study

Software	Purpose
Pix4Dmapper Version 4.4.12	Display and check the UAV orthomosaic project, convert .p4b (point cloud data) to .las
R 3.4.3 and R-studio Version 1.2.5042	Spectral reflectance extractions, statistical analysis, machine learning algorithms
ArcMap 10.7.1	Visualization, maps layout
LAStools	Height metrics derivation
Microsoft Word	Report and thesis writing
Microsoft Excel	Statistical analysis, graphs visualization
Microsoft Powerpoint	Thesis presentation

3. METHODOLOGY

This methodology section includes: a flowchart that provides an overview of the steps taken in the study; statistical analysis of field and spectral data; vegetation indices calculation; height extraction from DSM and height metrics derivation; and rice plant dry weight and height estimation using three machine learning algorithms to answer the objectives and research questions. The flowchart of the study is shown in Figure 8.

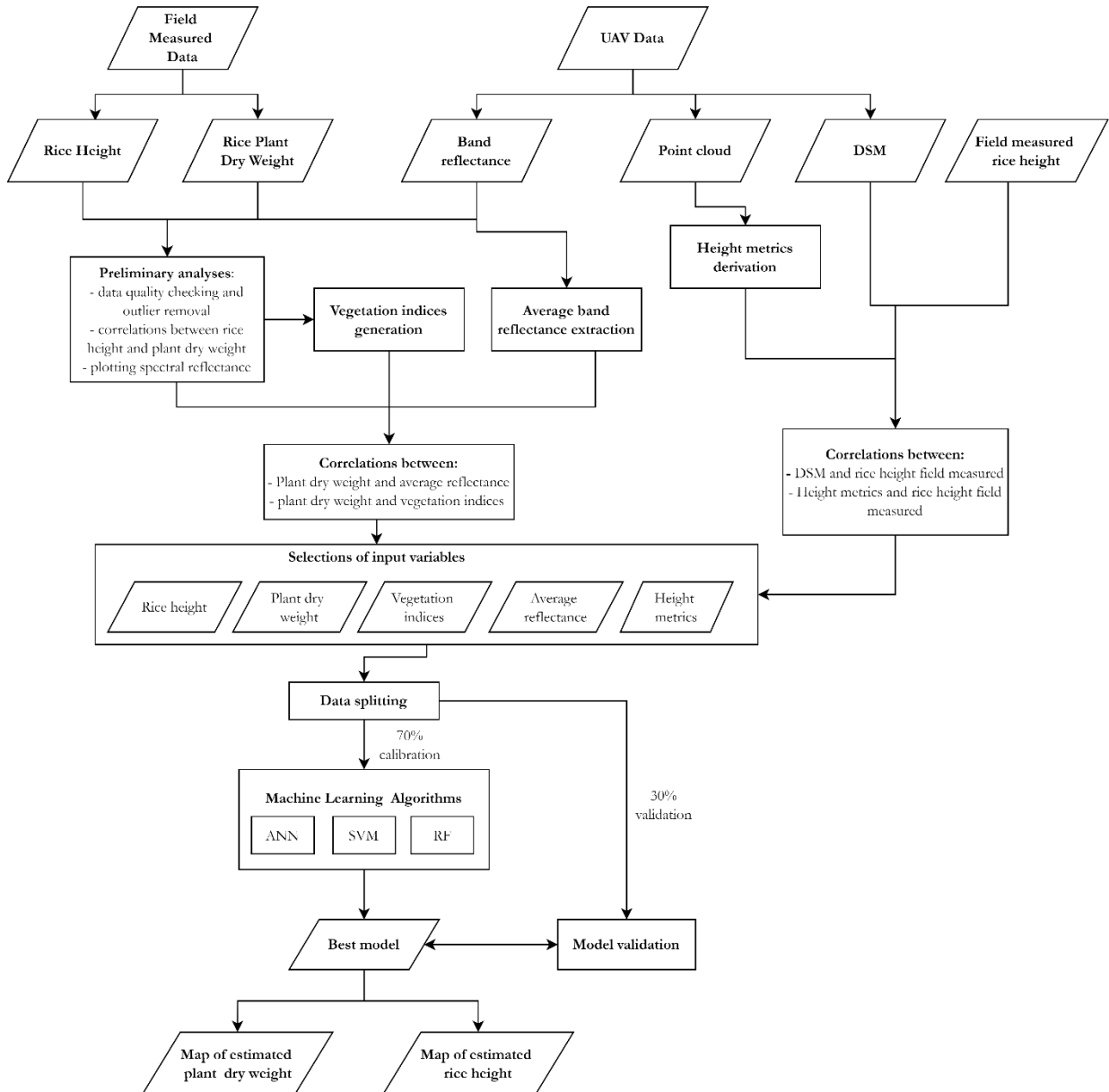


Figure 8. Flowchart of the study

3.1. Field data and spectral data variables exploration

This section presents the exploration of field data, consisting of total plant dry weight and rice crop height throughout the two seasons. The aim of this analysis was to understand the behaviour of the field and spectral data over time and to detect as well as to remove any possible outliers. The analyses were done by looking at the boxplots of each variable and at correlations between plant dry weight and rice height. The output of these data exploration are presented in Chapter 4.1 Preliminary Analyses Results.

Data exploration was also conducted for spectral data obtained from the multispectral UAV image. According to Curran (1980), multispectral remote sensing is capable of distinguishing and determining the state of features on the earth's surface, including vegetation amount such as dry weight. This is due to the fact the energy that reflected back -particularly from red and near-infrared- varies within different wavelengths, therefore vegetation features can be differentiated. In this study, the spectral data utilized consists of the reflectance of each multispectral band and vegetation indices derived from these bands.

The average reflectance of the subplots were extracted using a subplot boundaries shapefile, the the average reflectance was plotted in a graph in response to the spectral wavelength. The purpose of checking the extracted average reflectance and plotting them is to check the quality of the spectral reflectance data before using them in machine learning models, and to understand if there is any data that behaved unusually. Any unusual data were reported to IRRI for further investigation and advice. The plotting of spectral reflectance will be included as a preliminary result in Chapter 4.1.

3.2. Vegetation indices generation

Vegetation indices have been extensively utilized to estimate crop dry weight and crop height (Bendig, 2015; Prabhakara, Hively, & McCarty, 2015; Silleos, Alexandridis, Gitas, Silleos, & Alexandridis, 2008; Viljanen et al., 2018). In addition of including average spectral reflectance as predictors, we also computed vegetation indices as inputs to the machine learning algorithms. Six vegetation indices were calculated using R software. The six vegetation indices are:

- Ratio Vegetation Index (RVI), is one of the most commonly used vegetation index. It is calculated by simply dividing the spectral reflectance near infra-red band by the red band. The index was introduced by Jordan in 1969 with the basic principle that leaves absorbs more red than infrared light. Nowadays, due to a good correlation with plant dry weight and its sensitivity to vegetation, RVI widely used in monitoring and estimating dry weight.
- Normalized Difference Vegetation Index (NDVI), is another vegetation index that commonly used and first was introduced by Rouse, Haas, Schell, and Deering (1974) with the objective to produce a vegetation index that separates the green vegetation in the background from soil brightness. However, NDVI saturates at high dry weight values (Mutanga et al. 2004).
- Simple Ratio Red-Edge (SRRE), is similar to RVI, however it is a near-infrared band divided by the red edge band (Gitelson & Merzlyak, 1994).
- Normalized Difference Red Edge Index (NDRE), is usually used to detect healthy vegetation, as it is sensitive to chlorophyll content in leaves. NDRE is also less sensitive to saturation in dense vegetation, which can be occurred during the later stage of rice dry weight estimation (Bonfil, 2017).
- Soil-Adjusted Vegetation Index (SAVI), was developed to improve the sensitivity of NDVI to soil background (Huete, 1988). It is important to notice that the study area is an irrigated field, hence more water than soil.
- Modified Triangular Vegetation Index 2 (MTVI2), is one of the indices that is used to determine green Leaf Area Index (LAI) (Haboudane, Miller, Pattey, Zarco-tejada, & Strachan, 2004) but also has shown a good sensitivity at medium to high dry weight (Xiuliang Jin et al., 2015).

Table 6. Vegetation indices used in this study

Vegetation Indices	Formula	Reference
RVI	ρ_{NIR} / ρ_{Red}	Jordan (1969)
NDVI	$\rho_{NIR} - \rho_{RED} / \rho_{NIR} + \rho_{RED}$	Rouse et al. (1974)
SRRE	$\rho_{NIR} / \rho_{Red\ Edge}$	Gitelson and Merzlyak (1994)
NDRE	$\rho_{NIR} - \rho_{Red\ Edge} / \rho_{NIR} + \rho_{Red\ Edge}$	Gitelson and Merzlyak (1994)
SAVI	$\left(\rho_{NIR} - \rho_{Red} / \rho_{NIR} + \rho_{Red} + L \right) (1 + L)$	Huete (1988)
MTVI2	$\frac{1.5 \times [1.2(\rho_{NIR} - \rho_{Green}) - 2.5(\rho_{Red} - \rho_{Green})]}{\sqrt{(2\rho_{NIR} + 1)^2 - (6\rho_{NIR} - 5\sqrt{\rho_{Red}}) - 0.5}}$	Haboudane et al. (2004)

3.3. Statistics of field data and spectral data variables

The correlations between measured field data and mean spectral reflectance extracted from UAV images were studied. The aim of this assessment is to understand the relationship between field data and spectral data. Two relationships were established: correlations between total plant dry dry weight and average reflectance; and between total plant dry dry weight and vegetation indices. A strong correlation means that the variables are strong predictors. Strong predictors are expected to produce a model estimation with high accuracy.

3.4. Exploratory Analysis of Rice Height Data

Both DSM data and point cloud data were analysed and compared with field measured rice crop height to determine which if any had the best correlation with field measured height. If the correlations between DSM or point cloud data with field measured rice height are considered high, therefore one of the height dataset can be used as predictors to estimate rice height using machine learning algorithms. Another objective of obtaining rice crop height from DSM or point cloud data is whether one of these data can be predictors to improve dry weight estimations using machine learning algorithms. The DSM data was analysed by extracting the height information using the subplot boundaries shapefile, and the average height from each subplot was obtained. Then the correlation between height derived from DSM and field measured crop height data was made.

3.5. Machine Learning Algorithms

Three machine learning regressors (i.e., ANN, SVM, RF) were applied for the estimation of dry weight and then the performance of each was evaluated.

3.5.1. Artificial Neural Network (ANN)

ANN is one of the methods applied in this study to estimate dry weight and rice height. Many studies have conducted similar research using this method (Ali, Cawkwell, Dwyer, & Green, 2017; Devia et al., 2019; Han et al., 2019; Y. Q. Jin & Liu, 1997; Reisi-Gahrouei, Homayouni, McNairn, Hosseini, & Safari, 2019). In this study, the software used to run the ANN method was R with the *neuralnet* package (Fritsch, Guenther, Wright, Suling, & Mueller, 2019). The most commonly used model of a neuron is shown in Figure 9. The inputs $x_1 \dots x_n$ have connections which are multiplied by the weights $w_1 \dots w_n$. The sum of the weights then pass through an activation function, which is used to limit the output of the neuron (Mas & Flores, 2008)

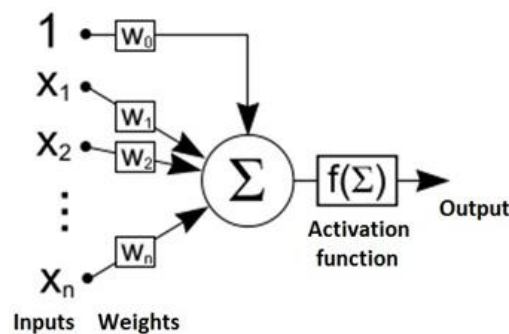


Figure 9. Anatomy of artificial neuron (Chalupník, 2012)

Since this is a supervised neural network, the dataset must be split into training and testing set. Splitting the dataset is necessary to create a training and testing dataset, and to avoid an overfitting model. Overfitting is when a model learns the details and noise in the training data too well, which then negatively impacts the model's performance. Training and testing sets were derived from the data based on a random selection using a 70:30 ratio. The training set contains the samples of the dataset used to fit the parameters of the model. In the process of training the data, it includes the input data as input layer, hidden layer, and output layer as the output of the model. This ANN architecture is commonly known as multilayer perceptrons (MLPs) (Figure 10). The neuralnet package is essentially a black box where user cannot determine much about the fitting, the weight, and the model. However, the user may define the number of neurons and the number of hidden layers. There are no specific rules to determine the hidden layer, it relies on user experience and experimentation (Skidmore, Turner, Brinkhof, & Knowles, 1997). After experimenting with several numbers of hidden layers, the ANN model in this study was set to have one hidden layer with three neurons. Then, the trained dataset was used to validate the model.

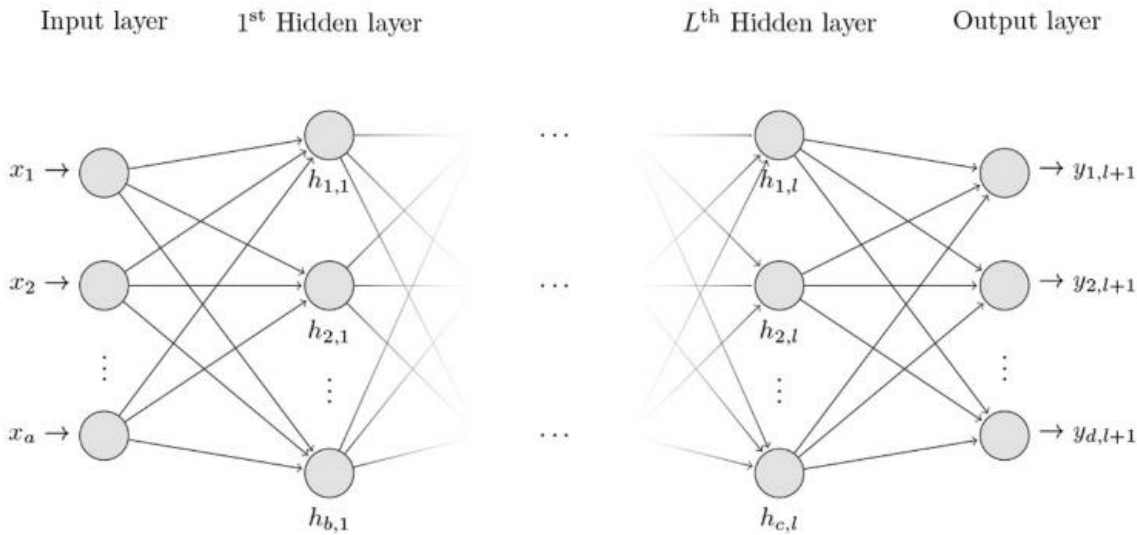


Figure 10. An example of the architecture of ANN-MLPs model (Castro, Oblitas, Santa-Cruz, & Avila-George, 2017)

3.5.2. Support Vector Machine (SVM)

The Support Vector Machine (SVM) for regression method used the R software with the *caret* package (Khun et al., 2020). SVM is one of the most widely used machine learning algorithms for estimating crop parameters, such as dry weight and height (Han et al., 2019; Karimi, Prasher, Madani, & Kim, 2008; Moeckel et al., 2018; Ndikumana et al., 2018). This algorithm was first introduced as a statistical learning theory by Vapnik (1995). The SVM algorithm focuses on minimizing the bounds on the risk function, rather than minimizing the error on the training data. The main feature of the SVM algorithm is the kernel function. The kernel function is used to map the non-linear observations into a higher dimensional space (Murty & Raghava, 2016). There are four types of kernel functions including linear, polynomial, RBF, and sigmoid kernels. The aim of SVM regression is to build a linear function that can make the closest approximation of the dependent variables (Karimi et al., 2008). In this study, the modelling with SVM regression was performed using the linear kernel method, as the number of output variables is smaller than the number of input variables (Hsu, Chang, & Lin, 2016). Other than choosing the SVM linear method, the other parameters of SVM regression that need to be chosen or tuned are the number of k-fold cross-validations and c (cost). The same data partitioning as used in ANN was also used in this method, using the 70:30 ratio of training and testing data. The modelling of SVM regression in this study used 10-fold cross validation repeated 5 times. As for the parameter c , also known as cost, in this study we tested 13 c values, starting from 0, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, and 5.00. The *caret* package will automatically pick the c value with smallest RMSE.

3.5.3. Random Forest (RF)

The Random Forest (RF) method used R software, with the *caret* package. The Random Forest algorithm is a combination of many decision trees, where each is trained independently, and the final prediction is made by averaging the individual trees (Breiman, 2001). RF is trained with bootstrap aggregating, or bagging, where subsets of training data are randomly sampled then use this smaller set of data to fit into the model, and aggregate the predictions. This is also called as tree bagging. The bagging method is also applied to the

feature space. It reduces the correlations between predictor variables by training them, instead of the entire set of independent variables or the feature, but on randomly sampled independent variables (Bryll, Gutierrez-Osuna, & Quek, 2003). There are some advantages in running the RF algorithm; on a large dataset it runs efficiently, has fewer parameters compared to ANN and SVM, it is not sensitive to noise, and it is able to handle hundreds or thousands variables without a need to remove some (Xiu liang Jin et al., 2013; Wang, Zhou, Zhu, Dong, & Guo, 2016). There are two important parameters in RF method, consisting of *mtry* and *ntree*. *Mtry* indicates the number of variables randomly sampled as candidates at each split and *ntree* indicates the number of trees to grow. The caret package will select *mtry* with the smallest RMSE. As for *ntree*, it was decided to use the default of 500 trees.

3.5.4. Machine learning generated models

In this study, there will be four different sets of input variables to run three machine learning algorithms. From hereon they will be named as Model 1, Model 2, Model 3, and Model 4. This was repeated for 2016 and 2017. The aim of producing the four models is to find out which combinations of input variables produce the most accurate model. The selections of the input variables were based on the preliminary analysis of the field data and spectral data variables (Section 3.5).

The dependent variable for the dry weight model is the total plant dry weight on DAT 62 for 2016 and DAT 68 for 2017 from field measurements, representing the measured dry weight closest to harvest date. As for the height model, the dependent variable is rice height from the single height measurement taken during the season on DAT 62 for 2016 EWS and DAT 68 for 2017 DS. The independent variables are the spectral bands and VIs from UAV data through the season and in the case of the dry weight model, measured dry weight earlier in the season. As mentioned earlier, before running the models, the dataset was divided into two sets, the training set and the testing set with a ratio of 70:30. These models were evaluated using the metrics explained in Section 3.8.

Model 1:

Model 1 used 39 variables as predictors, including the historical data of total plant dry weight throughout the season, the mean reflectance of red edge and near-infrared bands, and MTVI2 index data throughout the season. The dependent variable here is the total plant dry weight on DAT 62 for 2016 EWS model, and total plant dry weight in DAT 68 for 2017 model. Here the red edge and near infra-red bands were selected due to the fact both have high correlations to total plant dry weight. The same reason goes for the selection of MTVI2 due to its high correlation with plant dry weight. Both mean reflectance data and MTVI2 were included from 12 dates of UAV data acquisition.

Table 7. The list of independent variables as input for the model 1

		Amounts of variables
Independent variable	Historical total plant dry weight	3
	Mean reflectance of red edge and near infrared bands	2 bands × 12 dates (flights) 24
	MTVI2	12 dates (flights) 12
Total variables		39

Model 2:

Model 2 used only 16 predictors, including historical data of the total plant dry weight and MTVI2 index throughout each season (12).

Table 8. The list of independent variables as input into model 2

		Amounts of variables
Independent variable	Historical total plant dry weight	3
	MTVI2	12 dates (flights) 12
Total variables		16

Model 3:

Model 3 was using 63 variables as predictors, among them are the historical data of total plant dry weight, the temporal mean reflectance (12) of 4 bands of the multispectral image, and MTVI2 index data throughout each season (12).

Table 9. The list of independent variables as input to model 3

		Amounts of variables
Independent variable	Historical total plant dry weight	3
	Mean reflectance of green, red, red edge and near infrared bands	4 bands × 12 dates (flights) 48
	MTVI2	12 dates (flights) 12
Total variables		63

Model 4:

Model 4 used 60 variables as predictors, eliminating the historical data of total plant dry weight, and including the temporal mean reflectance (12) of 4 bands of the multispectral image, and MTVI2 index data throughout each season (12).

Table 10. The list of independent variables as input to model 4

Independent variable		Amounts of variables
Mean reflectance of green, red, red edge and near infrared bands	4 bands × 12 dates (flights)	48
MTVI2	12 dates (flights)	12
Total variables		60

3.5.5. Variable importance using RF regression algorithm

The aim of this analysis is to find which input variables is the best performer. Acknowledging the most important input variable that explains the variance of response variable will lead to easier understanding in building a model with high performance.

3.6. Model Evaluation

There are three metrics evaluation to evaluate model performance. Root mean square error (RMSE) is one of the most frequently used measurement to evaluate a model. It represents the sample standard deviation of the differences between the predicted and observed values. The formula of RMSE is as follow:

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

R^2 represents the coefficient of how well the predictive values fit compares to the field measured values. The metric is expressed:

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

Another metric to evaluate the model is Normalized RMSE (NRMSE). By normalizing RMSE, the model performance is comparable for analysing models with different units. The formula of NRMSE is as follow:

$$NRMSE = \frac{RMSE}{O}$$

Where y_i is the predicted value of dry weight and \hat{y}_i is the observed value of dry weight, meanwhile O is the range of observation data.

4. RESULTS

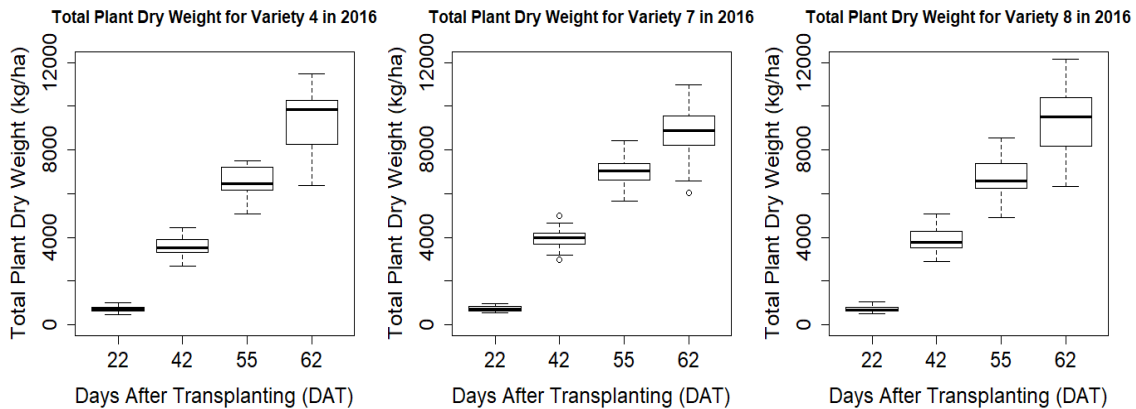
This chapter starts with the preliminary results from all analyses in preparing the input variables before going into the machine learning algorithms, then continues to the results of statistical analyses for variable selection, and then the machine learning process in estimating dry weight. After that, the results of the height data analysis is presented before the final maps.

4.1. Preliminary Analyses Results

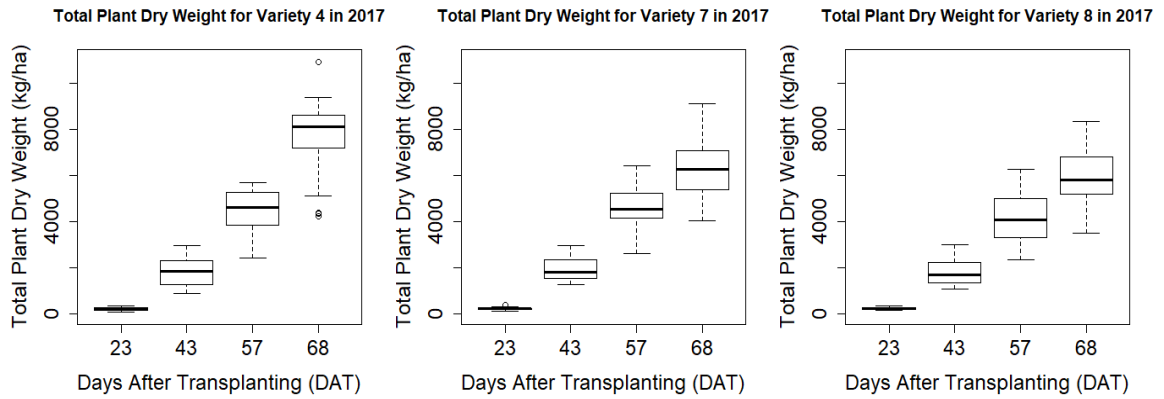
This section presents the preliminary results of field data and exploratory analyses of the spectral data.

4.1.1. Field Data Analysis

Figure 11 shows the data distribution as boxplots of total plant dry weight of the three rice varieties in each season. Figure 11 (a) and Figure 11 (b) show similar trends, in terms of the variability in plant dry weight. Towards the end of the season, the variability of plant dry weight values were increasing, from DAT 22 to DAT 62 and from DAT 23 and DAT 68. Figure 11(b) showing total plant dry weight for variety 8 in 2017, had two points of outliers in which were already removed by this stage after confirmation by Dr. Roland Buresh (personal communication with IRRI). These two outliers had very high values of total plant dry weight.

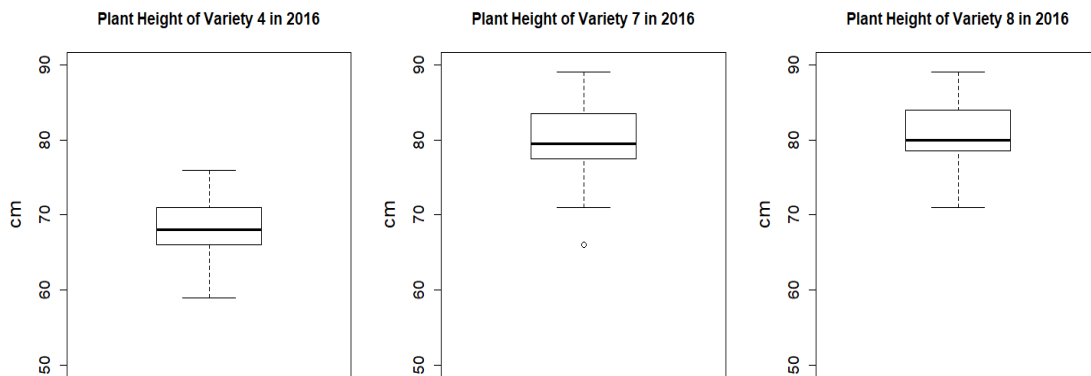


(a)

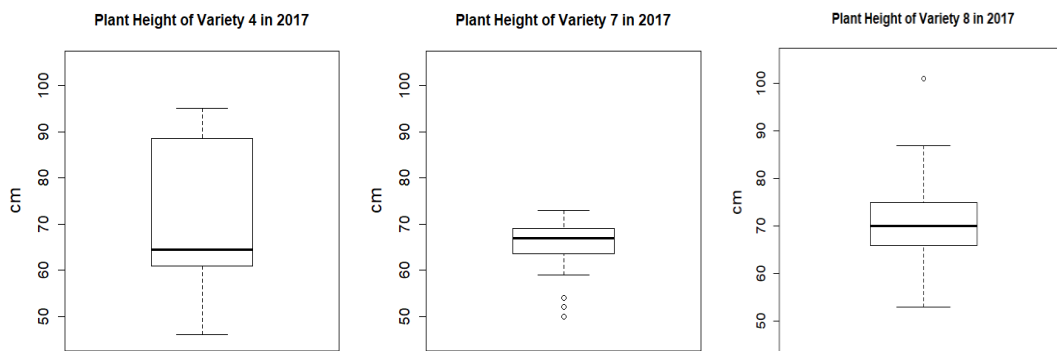


(b)

Figure 11. Boxplots showing the variation of total plant dry weight of rice at different dates (days after transplanting) during the growing season for three different varieties in the two seasons (a) 2016 EWS and (b) 2017 DS in the IRRI LTCCE.



(a)



(b)

Figure 12. Boxplot of rice crop height of the three different varieties in 2016 EWS (a) and 2017 DS (b).

Figure 12 show boxplots of rice crop height for variety 4, 7, and 8 in the year 2016 EWS (a) and year 2017 DS (b). The overall picture of these boxplots is that all varieties in both seasons show different distributions

of plant height as indicated by the median value and also the range of values in each box (indicating a high variability within the datasets). IRRRI measured the height by selecting only four plants and stretching the tallest leaf within one plot. This may explain the high variability of rice height in each plot and hence for each variety. The final number of observations for 2016 EWS is 72 and for 2017 DS is 70 points or samples.

In this section, Pearson's correlations were made between plant dry weight and plant height during the growing season. The correlations presented in Figure 13 for different varieties show similar trends where towards the end of the season, the correlations between plant dry weight and plant height are increasing. This means the most abundant dry weight occurred at the end of the season when the plant was the tallest. Based on this finding, rice plant height can be a reasonable predictor in estimating dry weight as they exhibit a strong positive correlation.

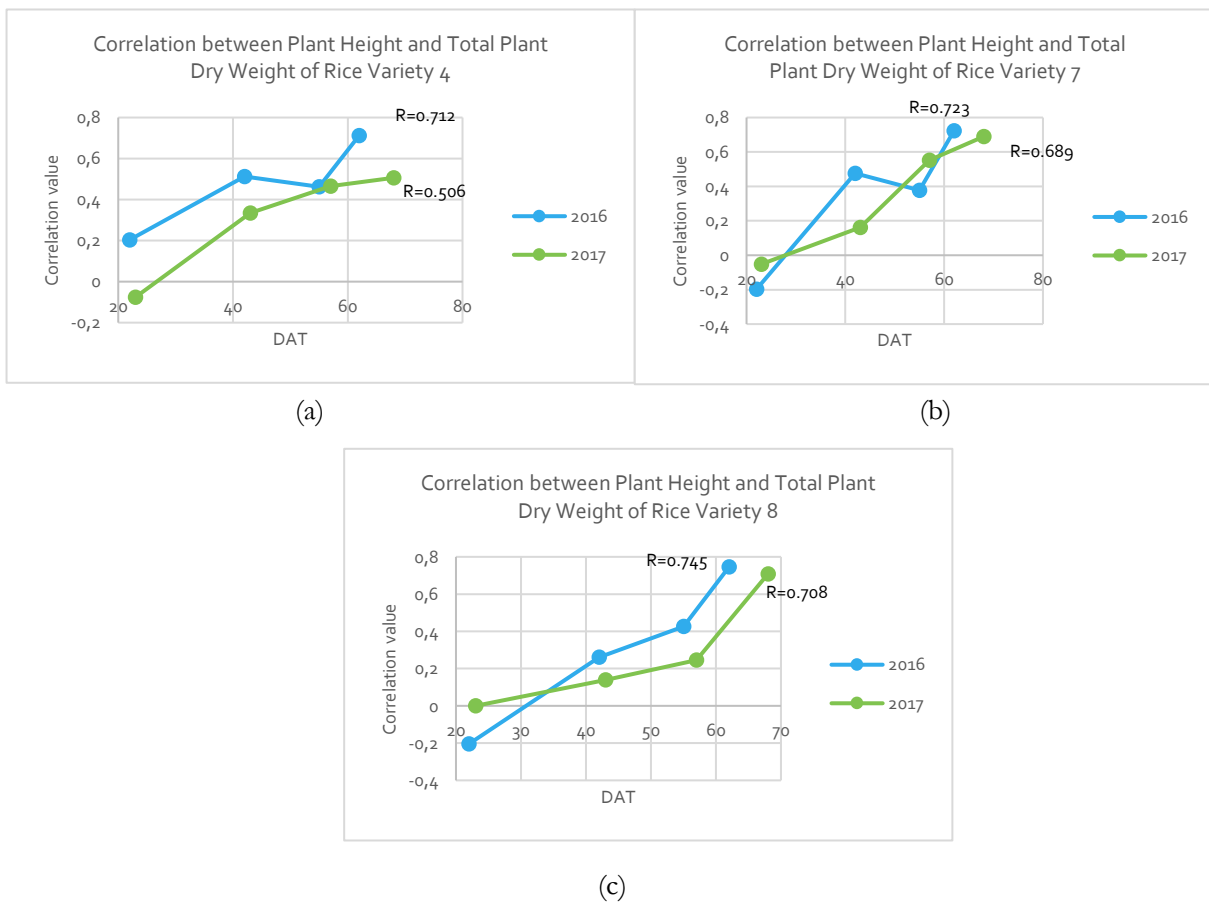
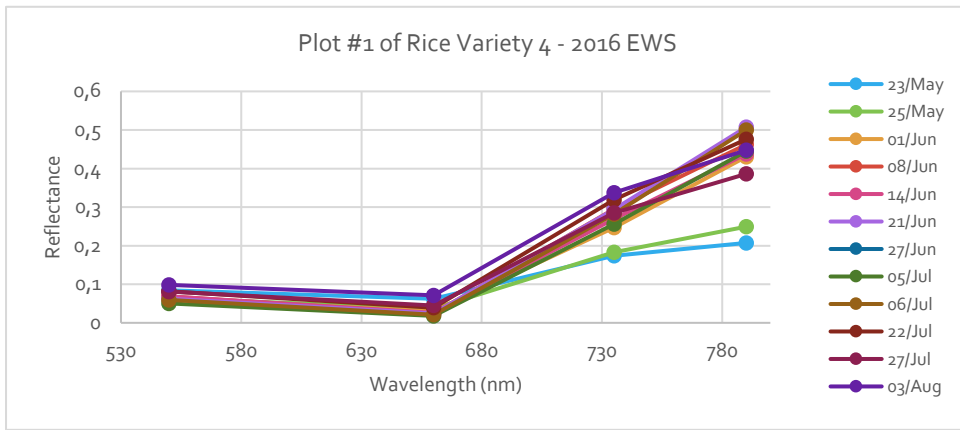


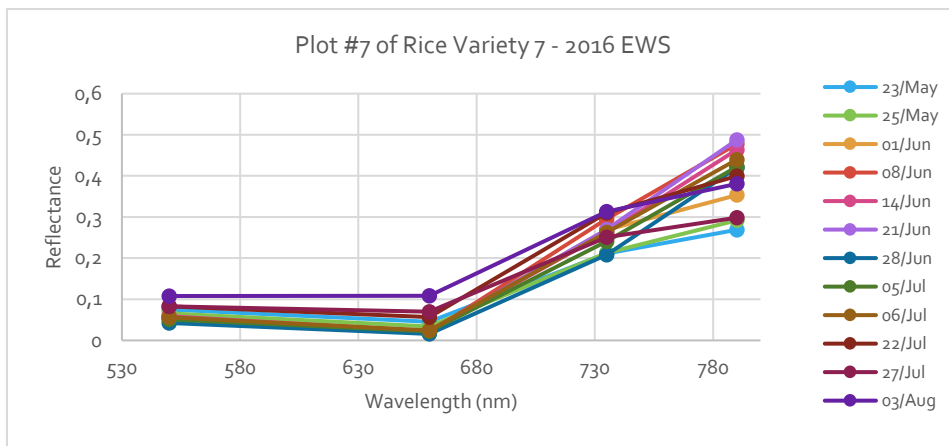
Figure 13. Correlation between plant height and total plant dry weight in the 2016 EWS and 2017 DS

4.1.2. Spectral Data Analysis

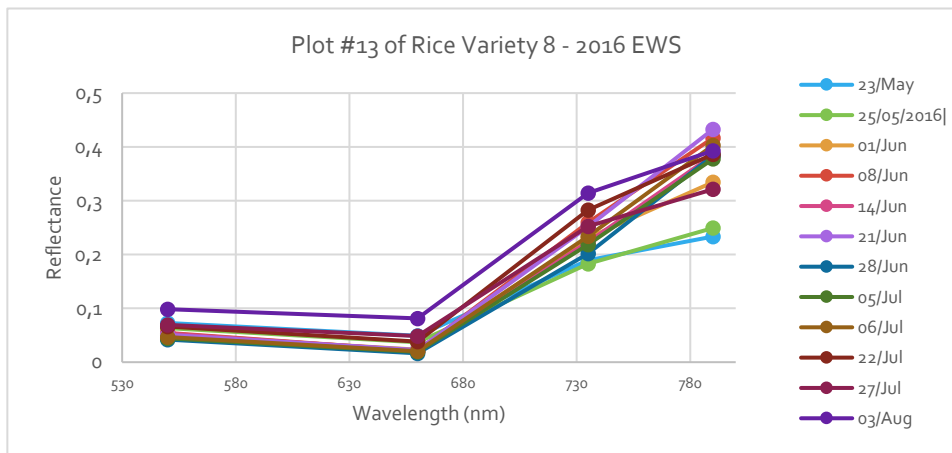
The average reflectance per date was plotted for four bands: green (550 nm), red (660 nm), red edge (735 nm), and near infrared (790 nm) as shown for three selected subplots in Figure 14 for all three rice varieties for 2016 EWS and in Figure 15 for 2017 DS. The plotting of spectral reflectance of both seasons show a similar trends. The red band has the lowest reflectance values due to absorption of chlorophyll, meanwhile the near infra-red band shows a peak in reflectance values due to the structure of vegetation leaves that makes the reflectance high. These two bands which show extreme responses of absorption and reflectance were a factor in the selection of vegetation indices for the machine learning algorithms.



(a)

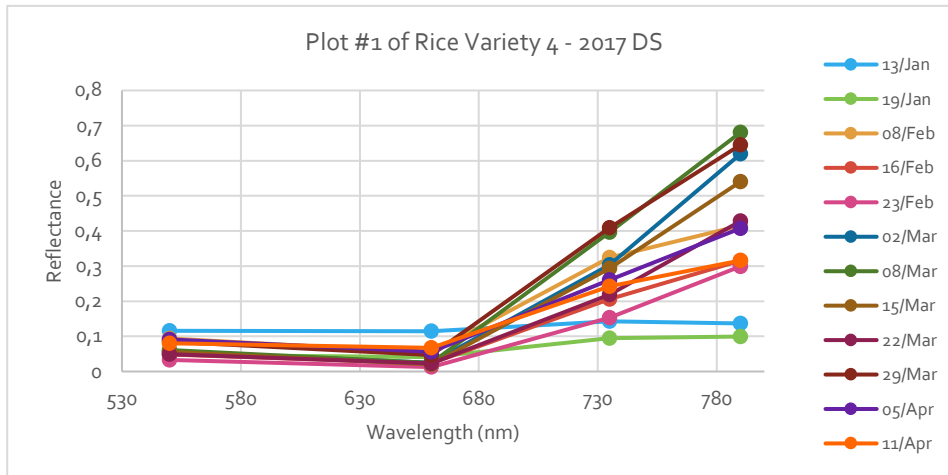


(b)

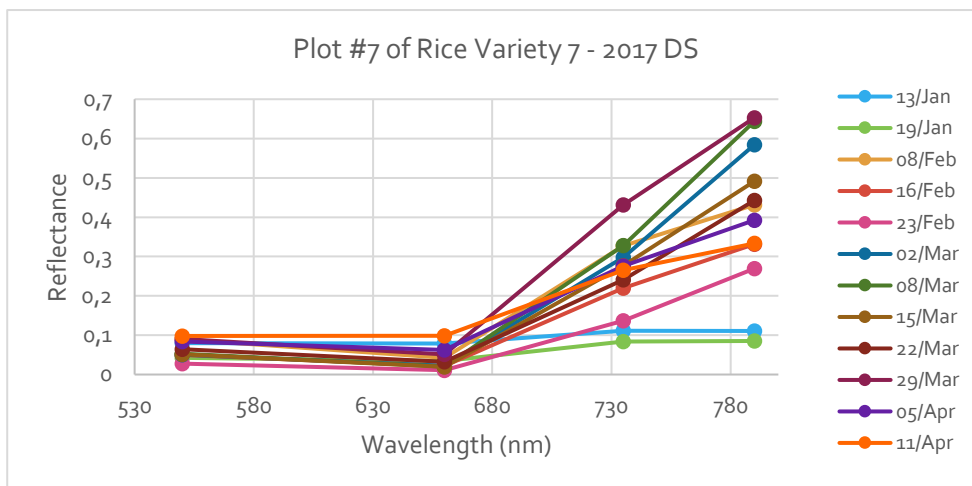


(c)

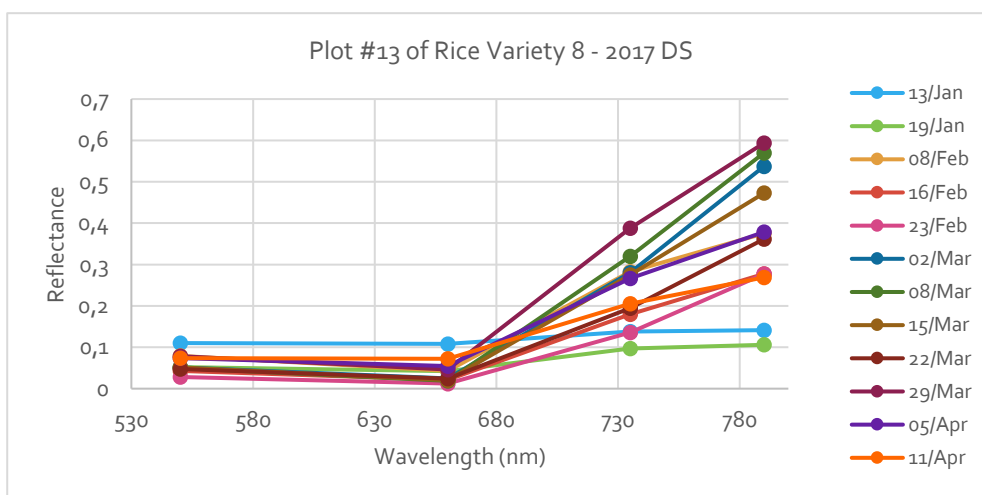
Figure 14. Mean reflectance of rice fields of green, red, red edge, and near infra-red bands throughout the 2016 EWS.



(a)



(b)



(c)

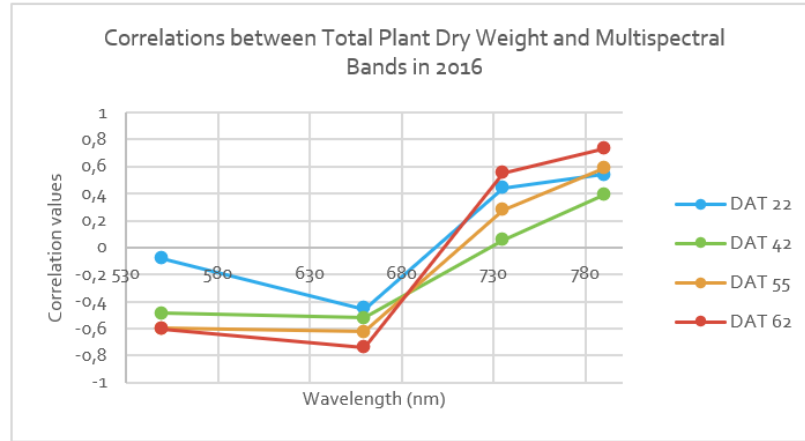
Figure 15. Mean reflectance of rice fields in green, red, red edge, and near infra-red bands throughout the 2017 DS

4.2. Statistics of Field Data and Spectral Data Related to Plant Dry Weight

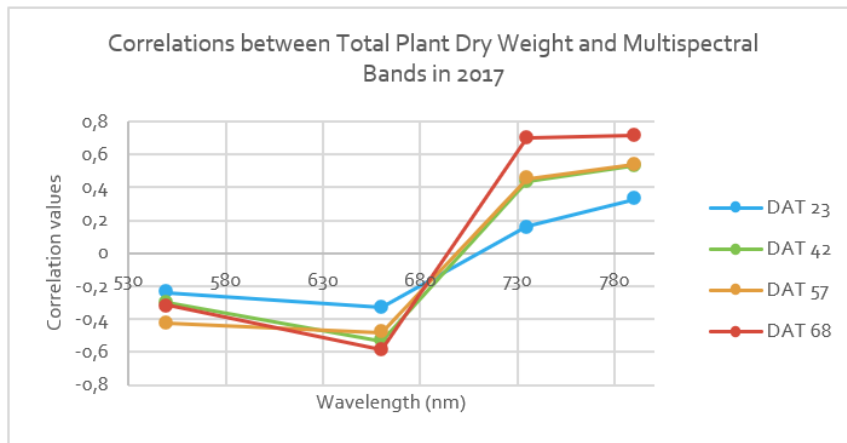
This section presents the correlations between field data variables and spectral data. There are two correlation analyses established between these two data, they are between total plant dry weight and average spectral reflectance, and between total plant dry weight and vegetation indices.

4.2.1. Correlations of Total Plant Dry Weight and Spectral Reflectance

Figure 16 shows the correlations of field measured total plant dry weight and spectral reflectance from UAV images in 2016 EWS and 2017 DS for all 72 field plots.



(a)



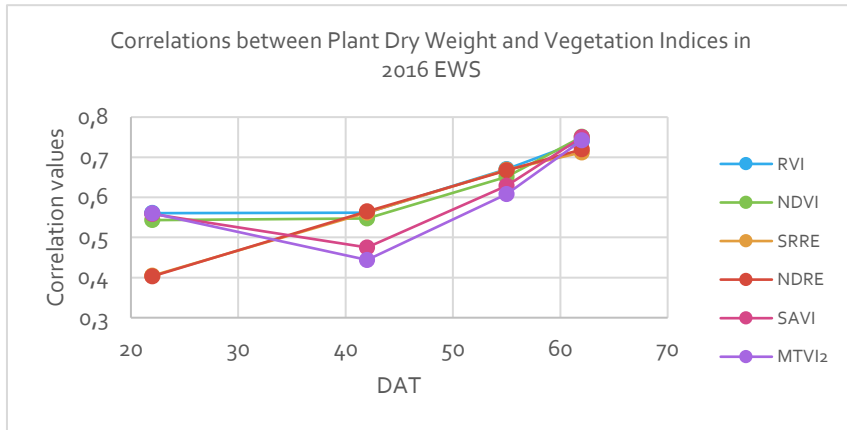
(b)

Figure 16. Correlations between total plant dry weight and multispectral bands in 2016 EWS (a) and 2017 DS (b)

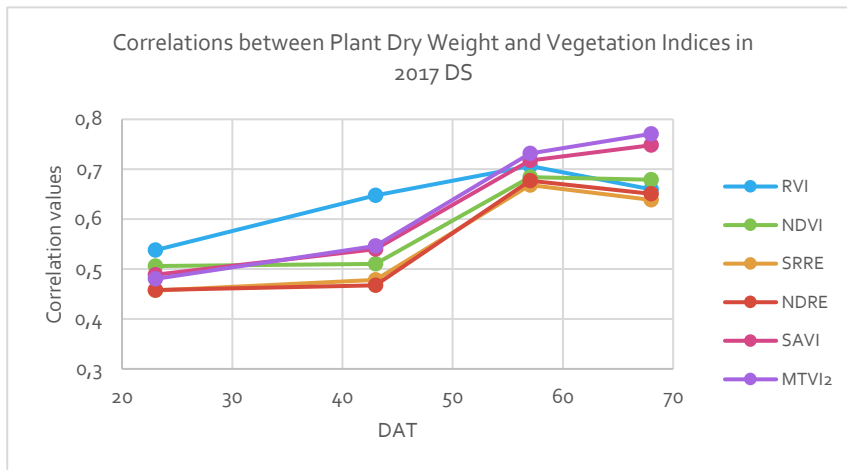
Based on Figure 16, it can be seen there are three bands that have strong relationships with the total plant dry weight or dry weight. Red band (660 nm) has a strong negative correlation, whereas near infra-red (790 nm) and red edge (735 nm) have a strong positive correlation. This finding will be considered for spectral band selection as input variables in the machine learning models.

4.2.2. Correlations of Total Plant Dry Weight and Vegetation Indices

After calculating six vegetation indices, correlations were established between the vegetation indices and total plant dry weight. Figure 17 shows the graphs of the correlations per season. Based on Figure 17 (a) and (b), the correlations between total plant dry weight and SAVI and MTVI2 have very similar trends and very close correlation values. MTVI2 was chosen as one of the input variables for the machine learning models.



(a)



(b)

Figure 17. The correlations of total plant dry weight and vegetation indices in 2016 EWS (a) and 2017 DS (b).

4.3. Machine Learning Models Results

This sub-section answers research question 1. There are four different set of data (models 1 to 4) for each machine learning algorithm for each each season. As mentioned in Section 3.6.4, four different models were produced by using a different set of input variables, and between these models the aim was to achieve the best model with the highest accuracy in terms of R^2 , RMSE, and NRMSE. This sub-section will present the model results per season.

2016 EWS

Model 1 used 40 input variables. There was one hidden layer containing three neurons in the ANN model and this setting was applied for all subsequent ANN models. The final c value used for the SVM model was 0.01 which brought the smallest RMSE value, and the final $mtry$ value for RF model was 26. Model 2 used 16 input variables with same input of hidden layer for ANN model. The final c value for SVM model was 0.01, and the final $mtry$ value for RF model was 11, which gave the smallest RMSE values. Model 3 with 64 input variables has the same c value as the other models with 0.01 for SVM model, and for RF model the $mtry$ value was 58. The summary of the parameters of each machine learning algorithm and its R^2 , RMSE and NRMSE values for this calibration or training phase is shown in Table 11.

Table 11. The summary of the calibration results for four models with 2016 EWS dataset

Model	ANN			SVM				RF			
	RMSE (kg/ha)	NRMSE (%)	R^2	Cost	RMSE (kg/ha)	NRMSE (%)	R^2	$mtry$	RMSE (kg/ha)	NRMSE (%)	R^2
1	1035	21.5	0.44	0.01	930	15.2	0.59	26	928	15.2	0.56
2	1476	18.1	0.32	0.01	919	15.0	0.59	11	929	15.2	0.57
3	982	16.8	0.46	0.01	1009	16.5	0.54	58	907	14.9	0.63
4	968	17.4	0.48	0.1	929	15.2	0.59	40	899	14.7	0.57

Next the result using testing data for validation are presented in the form of graphs in Figure 18 for Model 1, Figure 19 for Model 2, Figure 20 Model 3, and Figure 21 for Model 4.

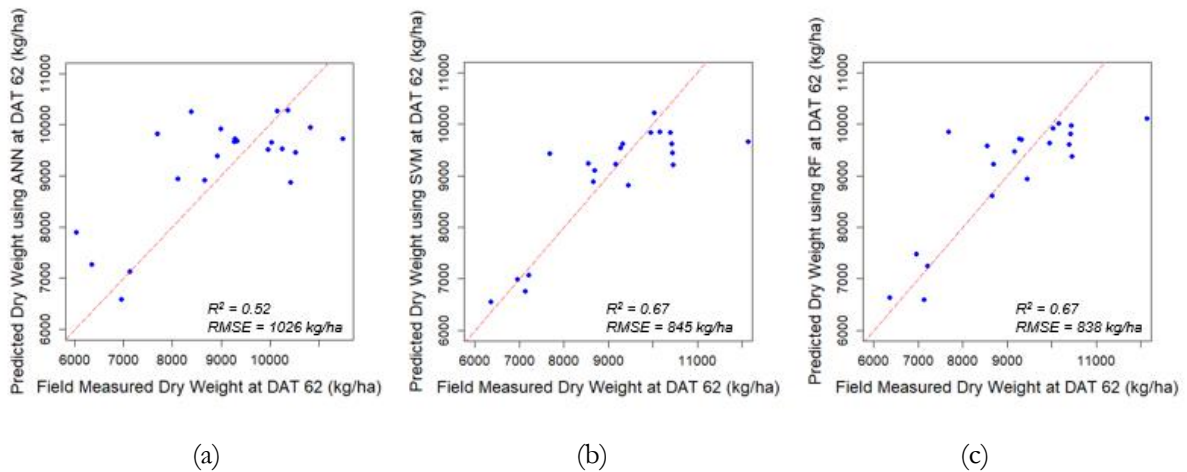


Figure 18. Model 1 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 data with their R^2 and RMSE. Red dotted line is the 1:1 line.

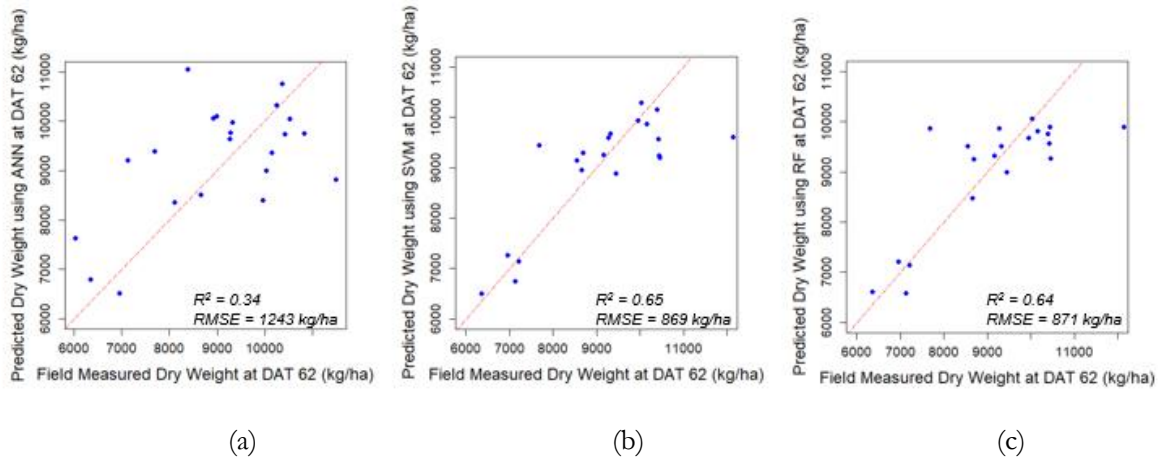


Figure 19. Model 2 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 data with their R^2 and RMSE. Red dotted line is the 1:1 line.

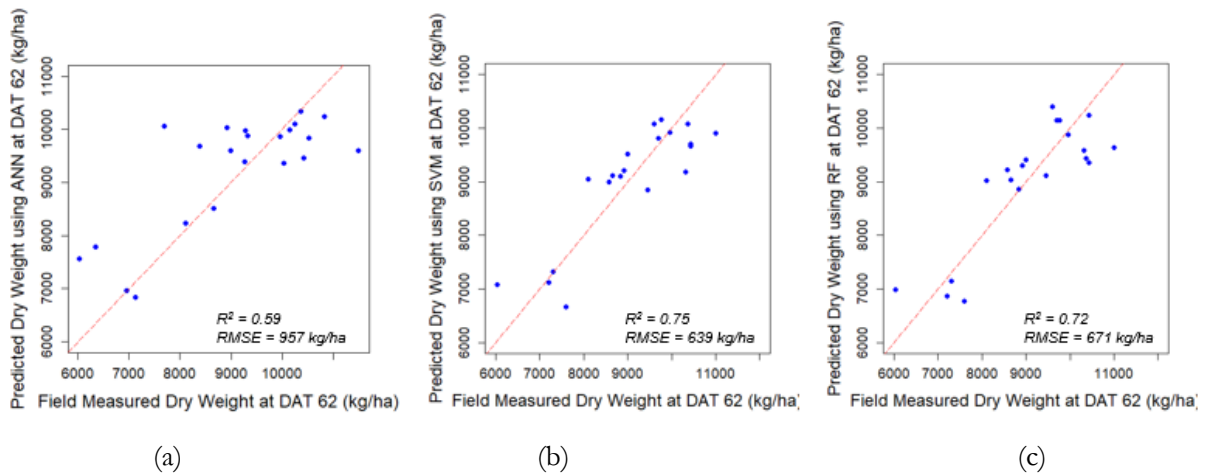


Figure 20. Model 3 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 EWS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

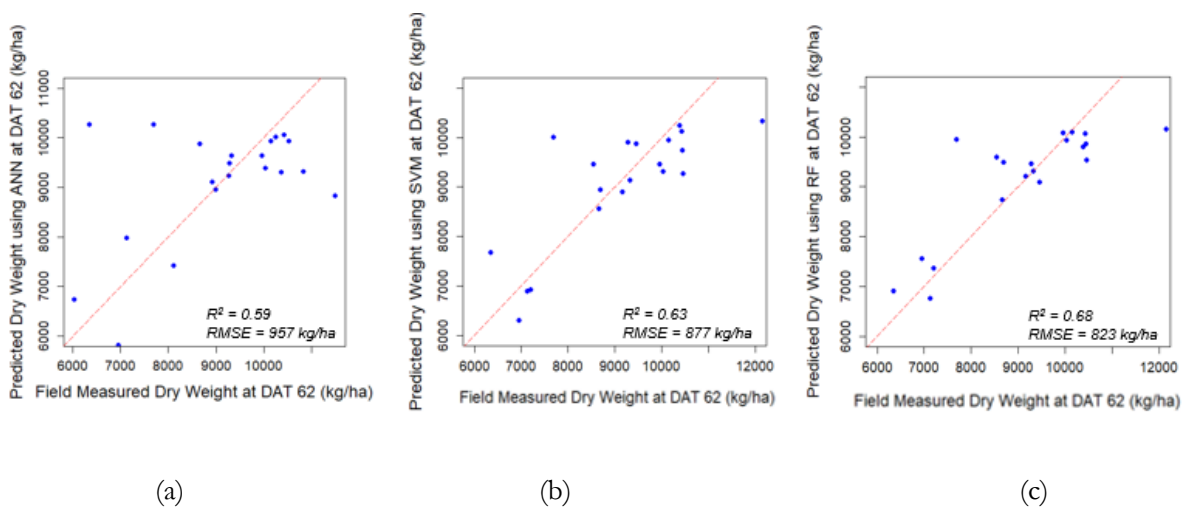


Figure 21. Model 4 results for predicted dry weight using ANN (a), SVM (b), and RF (c) for 2016 EWS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

The summary of the testing dataset is shown in Table 12, together with the training dataset for comparison.

Table 12. The summary results of ML algorithms when calibration and validation datasets of 2016 EWS were used.

Metrics	ML model	Calibration				Validation			
		Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
RMSE (kg/ha)	ANN	1035	1476	982	968	1026	1243	957	957
	SVM	930	919	1009	929	845	869	639	877
	RF	928	929	907	899	838	871	671	823
NRMSE (%)	ANN	21.5	18.1	16.8	17.4	20.3	16.8	15.6	15.6
	SVM	15.2	15.0	16.5	15.2	14.2	13.8	10.4	14.3
	RF	15.2	15.2	14.8	14.7	14.2	13.7	11.0	13.5
R ²	ANN	0.44	0.32	0.46	0.48	0.52	0.34	0.59	0.59
	SVM	0.59	0.59	0.53	0.58	0.67	0.65	0.75	0.63
	RF	0.56	0.57	0.63	0.57	0.67	0.64	0.72	0.68

Based on the result above, we can clearly see that within the 2016 EWS dataset, the SVM regression method shows the best result compared to ANN and RF in all three different sets of input variables, even though the predicted models of RF have very close evaluation metrics values to SVM regression method. Overall, SVM regression with Model 3 as input variables produced the best model (the red-highlighted result).

2017 DS

The input variables for the models in 2017 DS were kept the same as 2016 EWS. Model 1 with 40 variables had a single hidden layer with three neurons in the ANN, which was kept the same for all subsequent models. SVM had a c value of 0.05 and RF had an mtry value of 5. For Model 2 with 16 variables, SVM had a c value of 0.25 which gave the smallest RMSE value, and RF had an mtry of 2. Model 3 with 64 variables had a c value of 0.05 for SVM and an mtry value of 11 for RF. The summary result of the calibration dataset is presented in Table 13.

Table 13. The summary of the calibration results for four models with the 2017 DS dataset.

Model	ANN			SVM				RF			
	RMSE (kg/ha)	NRMSE (%)	R ²	Cost	RMSE (kg/ha)	NRMSE (%)	R ²	mtry	RMSE (kg/ha)	NRMSE (%)	R ²
1	1178	17.5	0.67	0.05	835	11.3	0.78	5	872	11.8	0.77
2	1365	15.2	0.56	0.05	878	11.9	0.76	2	942	12.7	0.69
3	1594	21.5	0.39	0.05	759	10.2	0.82	11	886	12.0	0.76
4	1477	18.4	0.54	0.05	725	9.8	0.82	5	849	11.5	0.78

The result using testing data for validation are presented in the form of graphs in Figure 22 for Model 1, Figure 23 for Model 2, Figure 24 for Model 3, and Figure 25 for Model 4.

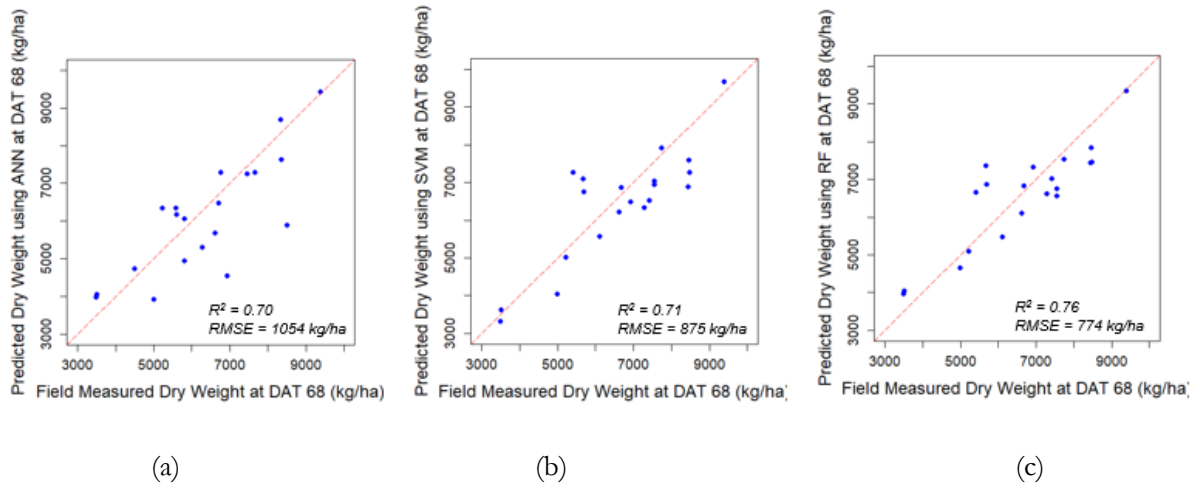


Figure 22. Model 1 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

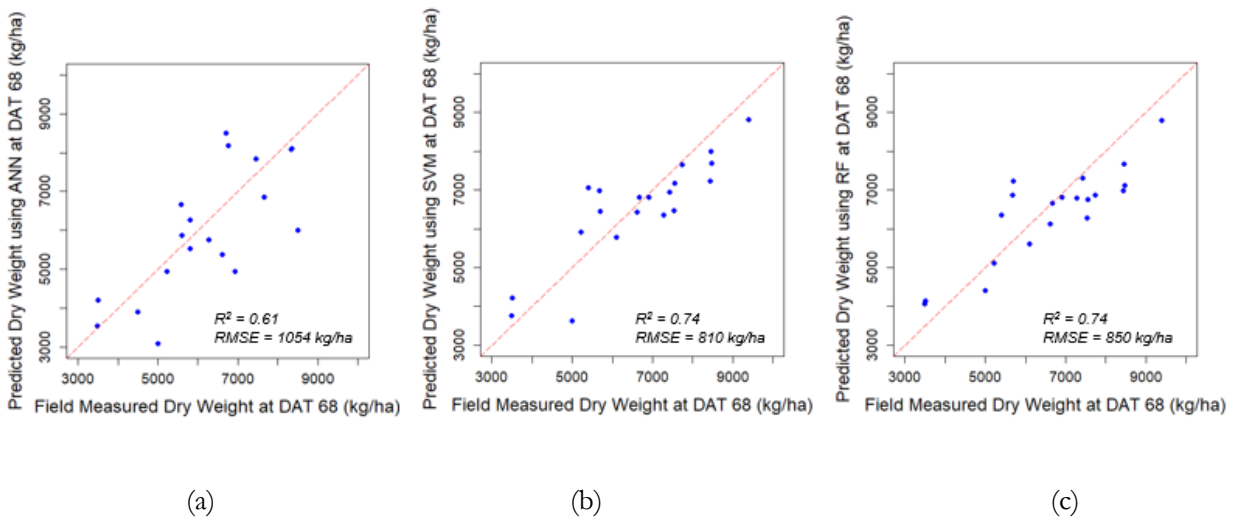


Figure 23. Model 2 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

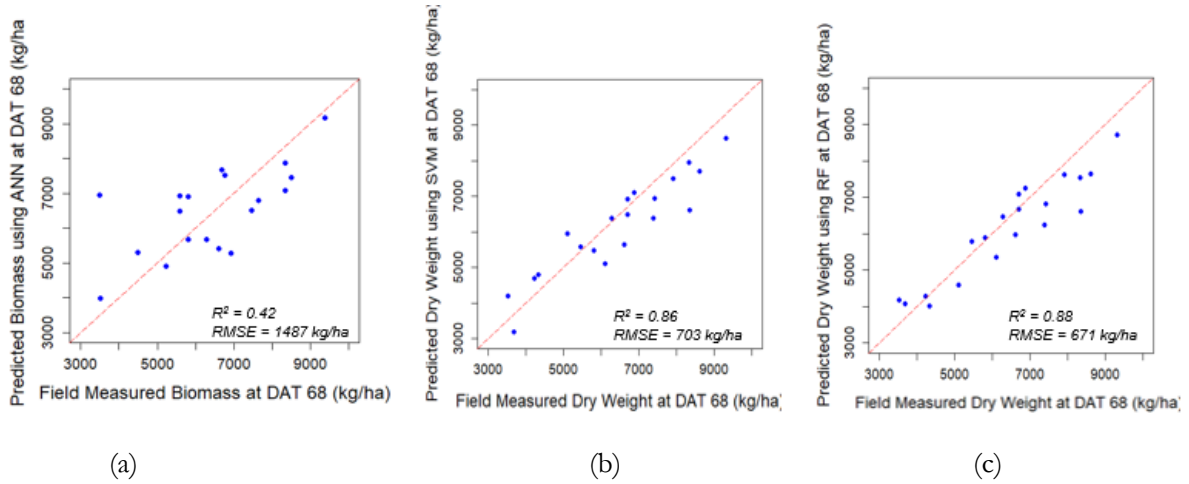


Figure 24. Model 3 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

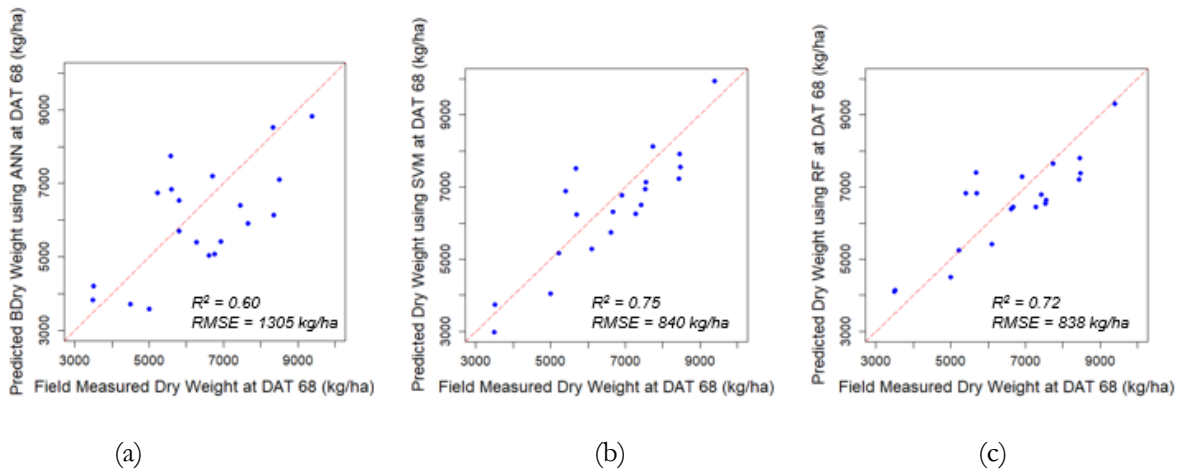


Figure 25. Model 4 results for predicting dry weight using ANN (a), SVM (b), and RF (c) for 2017 DS data with their R^2 and RMSE. Red dotted line is the 1:1 line.

The summary of the testing dataset is shown in Table 14, together with the training dataset for comparison.

Table 14. The summary results of ML algorithms when calibration and validation datasets of 2017 DS were used.

Metrics	ML model	Calibration				Validation			
		Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
RMSE (kg/ha)	ANN	1178	1365	1594	1477	1054	1263	1487	1305
	SVM	835	878	759	725	875	810	703	840
	RF	872	942	886	849	774	850	671	838
NRMSE (%)	ANN	17.5	15.2	21.5	18.4	17.0	14.2	20.1	17.6
	SVM	11.3	11.9	10.2	9.8	10.9	11.8	9.5	11.3
	RF	11.8	12.7	12.0	11.5	10.5	10.4	9.1	11.3
R ²	ANN	0.67	0.56	0.39	0.54	0.70	0.61	0.42	0.60
	SVM	0.78	0.76	0.82	0.82	0.71	0.74	0.86	0.75
	RF	0.77	0.69	0.76	0.78	0.76	0.74	0.88	0.72

The results of modelling four different sets of input variables for 2017 DS dataset show that RF produced better models, in comparison to ANN and SVM in terms of R², RMSE, and NRMSE. The NRMSE of RF method in four models show the lowest values, meaning better estimates of the variables. In addition, Model 3 with 64 variables also comes up as the set of variables with the highest R² and lowest NRMSE and RMSE values for 2017 DS dataset.

4.4. Variable Importance

After building models with selecting input variables based on the correlations between field measured data and spectral data, the establishment of variable importance using RF regression algorithm allows us to find which input variable performs the best. Table 15 shows the list of the variable importance of 2016 EWS model 3 which produce the best model. Increase in node purity (IncNodePurity) is measured by the reduction in sum of squared errors whenever a variable is chosen to be split.

Table 15. The list of variable importance of 2016 EWS data

Code of Variables	Importance (IncNodePurity)	Variables
mtvi2_5	100	MTVI2 obtained at 14 June 2016 (DAT 42)
re8	87.9	average reflectance of red edge at 5 July 2016 (DAT 62)
nir8	82.16	average reflectance of near infrared at 5 July 2016 (DAT 62)
nir12	59.19	average reflectance of near infrared at 3 August 2016
nir5	58.19	average reflectance of near infrared at 14 June 2016 (DAT 42)
nir6	57.77	average reflectance of near infrared at 21 June 2016
re12	57.38	average reflectance of red edge at 5 July 2016 (DAT 62)

mtvi2_8	54.83	MTVI2 obtained at 5 July 2016
nir7	54.38	average reflectance of near infrared at 28 June 2016 (DAT 55)
mtvi2_4	51.57	MTVI2 obtained at 8 June 2016
nir4	51.33	average reflectance of near infrared at 8 June 2016
re9	50.37	average reflectance of red edge at 6 July 2016
mtvi2_12	49.93	MTVI2 obtained at 3 August 2016
re6	45.85	average reflectance of red edge at 21 June 2016
mtvi2_7	43.82	MTVI2 obtained at 28 June 2016 (DAT 55)
r2	42.3	average reflectance of red at 25 May 2016
re10	41.56	average reflectance of red edge at 22 July 2016
re11	39.93	average reflectance of red edge at 27 July 2016
r4	27.78	average reflectance of red at 8 June 2016
nir10	26.21	average reflectance of near infrared at 22 July 2016

4.5. Extraction of Digital Surface Model (DSM) for plant height estimation

In this step, the average height data of the DSM image was extracted using the same plot boundaries shapefile during spectral reflectance extraction. A scatterplot of the average values of DSM image and the height data from field measurements is shown in Figure 26.

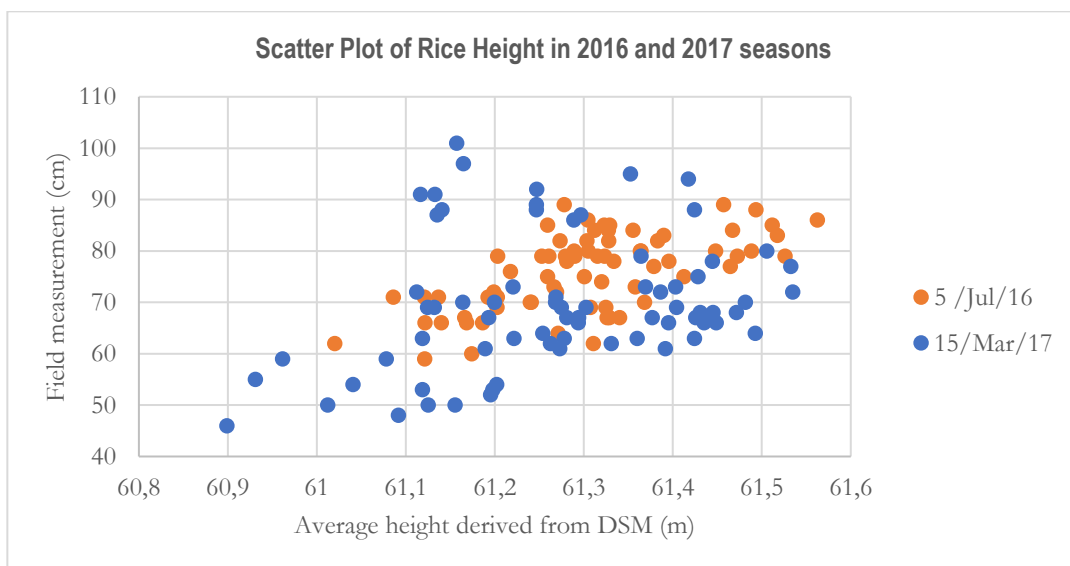


Figure 26. Field measured height vs height derived using DSM data (average value).

The correlation value between the height from field measurement and the height information extracted from DSM in year 2016 EWS was 0.61 meanwhile in the year 2017 DS the correlation value was very low, 0.27. Based on the scatterplot, it can be seen that height information from field measurements has high variability – from around 50cm to 100cm compared to the height information extracted from DSM – around 61cm.

4.6. Derivation of height metrics

After analysing the DSM data, height metrics were derived from the point cloud to see if there was greater variability in height information that could be extracted and compared to the field measured height. For each sample (subplot), 12 different height metrics were calculated (Table 16). The selection of these metrics was based on studies of Moeckel et al. (2018) and Viljanen et al. (2018). The height metrics derivation from point cloud data was to see if they provided a better representation of field measured height. The advantage of deriving height metrics compare to commonly single metrics method is that the information derived about crop height are different in each single metric (Moeckel et al., 2018). Moeckel et al. (2018) used height metric method to estimate the height of eggplant, tomato, and cabbage, meanwhile Viljanen et al. (2018) applied height metric method to estimate the height of grass. Both studies estimated crops height that considered low, therefore this study applied the same method to estimate rice height.

Table 16. Height metrics derived from point cloud data

Metric	Description
min	Minimum rice height
max	Maximum rice height
avg	Average rice height
std	Standard deviation of rice height
ske	Skewness of rice height
kur	Kurtosis of rice height
p25	25 th percentile of rice height
p50	50 th percentile of rice height
p70	70 th percentile of rice height
p90	90 th percentile of rice height
p95	95 th percentile of rice height
p99	99 th percentile of rice height

Table 17 shows that there was a low correlation between these height metrics derived from the point cloud to rice height measured in the field. Based on this finding, both sources of remotely sensed crop height have low correlations with field measured height. These weak correlations show that they cannot be used to estimate plant height as measured in the field using the field protocol described in Section 3.3, resulting in a negative finding for research question 2. The reasons for this negative finding will be discussed in section 5.

Table 17. Correlations between point cloud metrics and height obtained from field measurements

Correlation to height field measured	
min	0.00
max	0.12
avg	0.15
std	0.00
ske	0.06
kur	0.15
p25	0.16
p50	0.15
p75	0.14
p90	0.13
p95	0.12
p99	0.16

However, based on the relationship between field measured total plant dry weight and field measured height (Section 4.1), we introduced the height metrics into the plant dry weight model to see if it added any additional explanatory power to the models.

Using the best model from 2016 season, model 3 with SVM methods, the new model, with height metrics, produced a lower accuracy. The R^2 was 0.57 with RMSE of 867 kg/ha compare to the total plant dry weight model 3 with $R^2 = 0.75$ and RMSE of 630 kg/ha using 2016 EWS data. Hence, adding height metrics into the dry weight model failed to improve the model's performance. No further analysis was done with the height data.

4.7. Mapping Estimated Dry weight

To deliver the research objective 3, this section will present maps of field measured dry weight compared to estimated dry weight using the best machine learning algorithms in terms of R^2 and RMSE for season 2016 and 2017. Figure 27 displays the map with 72 field subplots of field measured dry weight compared to estimated dry weight using SVM algorithm in 2016. Figure 28 shows the difference map between field measured dry weight and the predicted dry weight in 2016 on the left side, and the percentage map on the right side. Meanwhile, Figure 29 displays the map with 70 field subplots of field measured dry weight compare to 72 field subplot estimated dry weight using RF algorithm in 2017. There were fewer field plots in 2017 season because two of the outliers in the total plant dry weight were removed. However, using the model, it is able to estimate dry weight on two missing subplots, but not on the difference and percentage difference map of 2017 (Figure 30). Both maps need the actual dry weight of two missing plots, therefore the difference and percentage difference of two missing subplot cannot be produced.

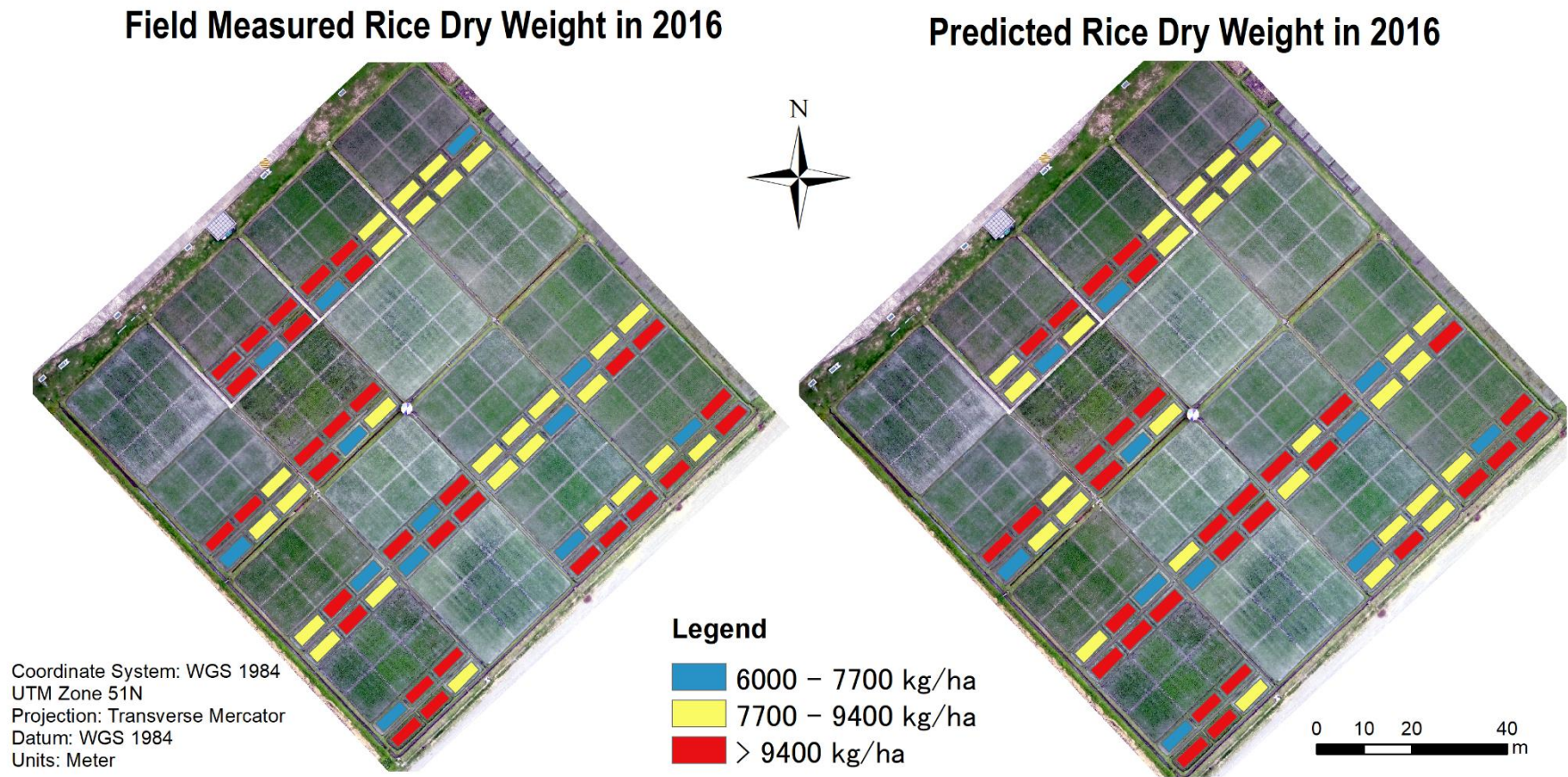


Figure 27. The maps of field measured dry weight and estimated dry weight using SVM algorithm in 2016

The difference map of field measured and predicted dry weight in 2016

The percentage difference map of field measured and predicted dry weight in 2016

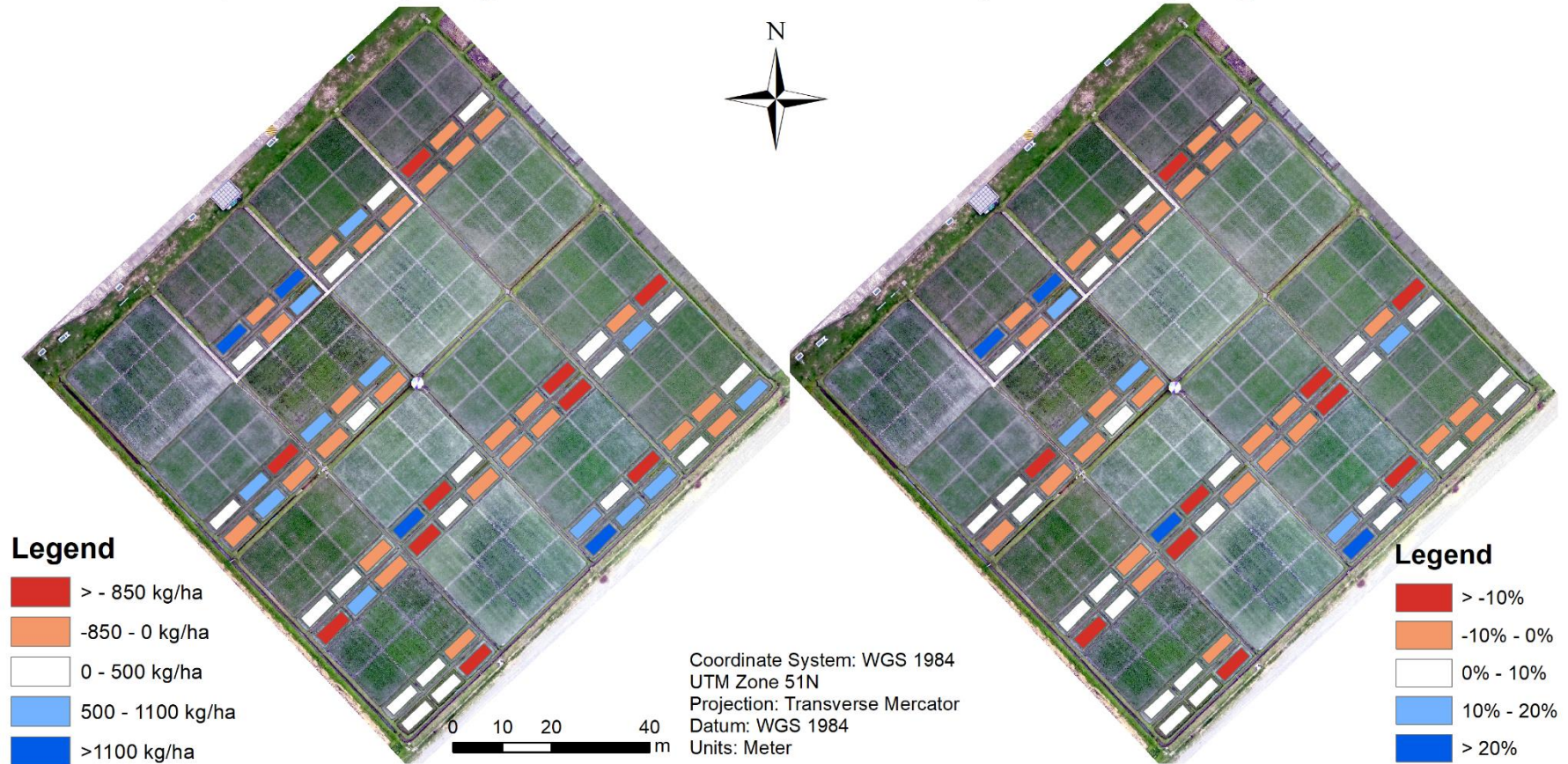
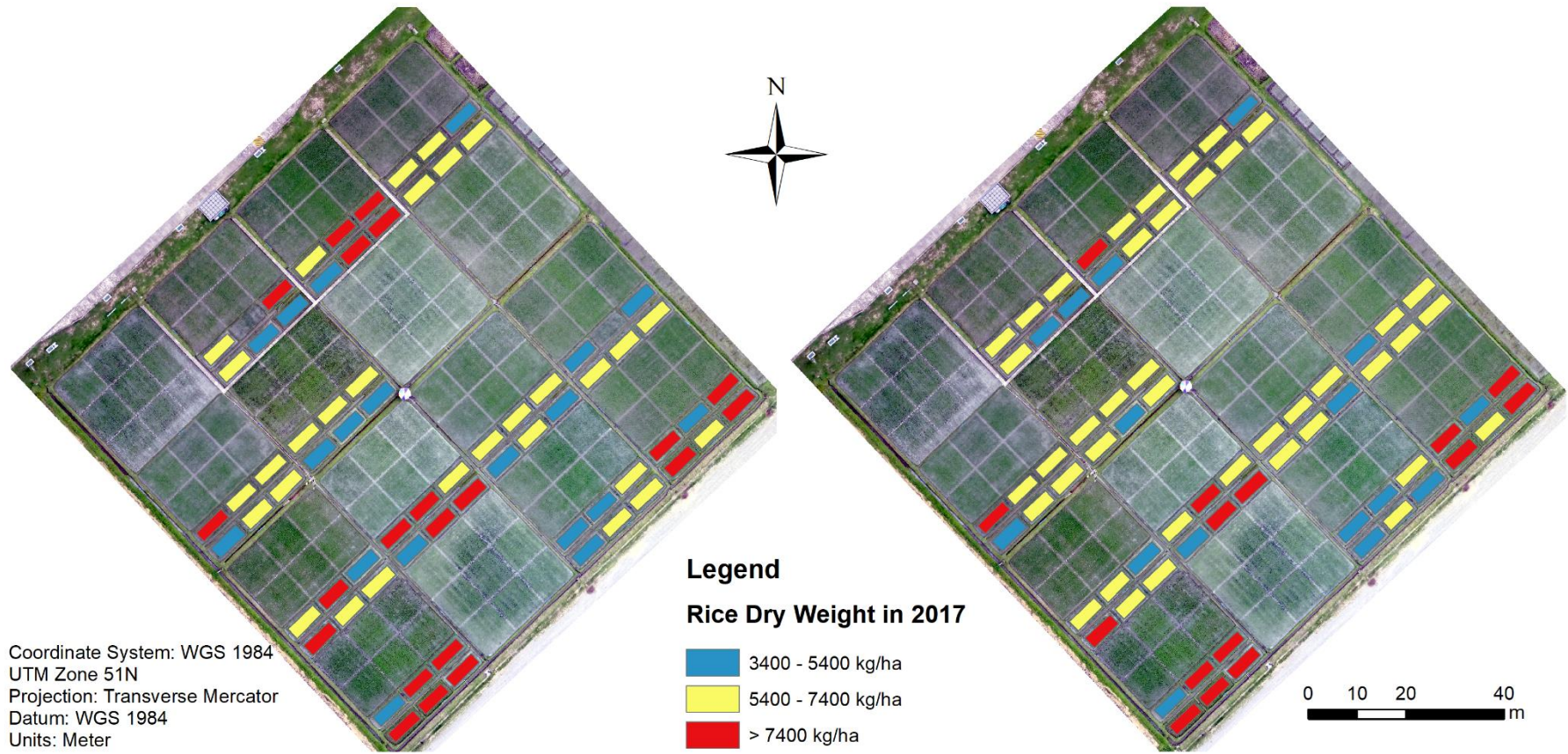


Figure 28. The difference map (field measured – predicted) of dry weight in 2016 (left) and the percentage difference map ($100 * (\text{field measured} - \text{predicted}) / \text{predicted}$) (right)

Field Measured Rice Dry Weight in 2017

Predicted Rice Dry Weight in 2017



Coordinate System: WGS 1984
UTM Zone 51N
Projection: Transverse Mercator
Datum: WGS 1984
Units: Meter

Figure 29. The maps of field measured dry weight and estimated dry weight using RF algorithm in 2017

The difference map of field measured and predicted dry weight in 2017

The percentage difference map of field measured and predicted dry weight in 2017

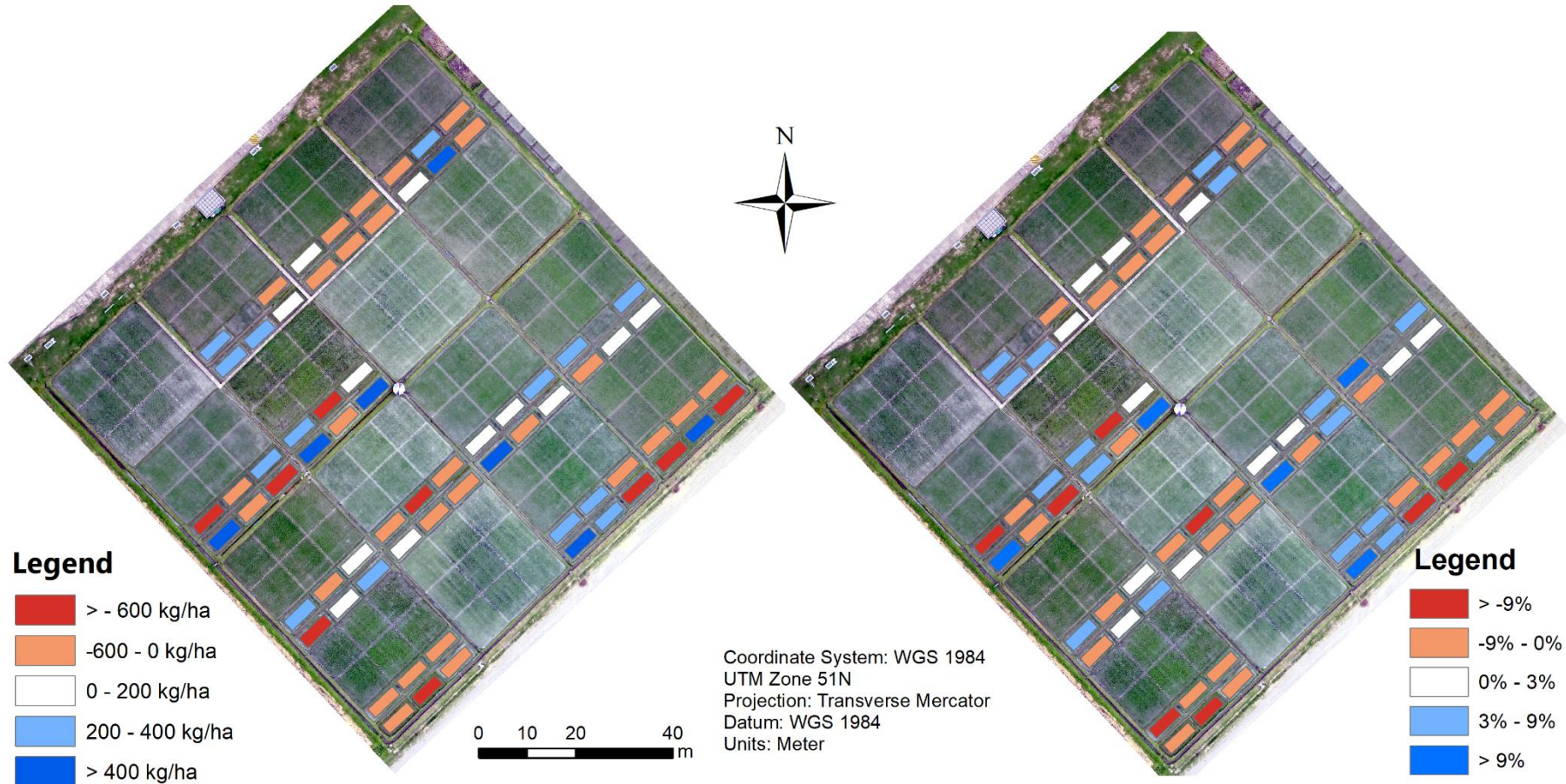


Figure 30. The difference map (field measured – predicted) of dry weight in 2017 (left) and the percentage difference map ($100 * (\text{field measured} - \text{predicted}) / \text{predicted}$) (right)

5. DISCUSSIONS

The study aim was to estimate and map plant dry weight and rice crop height using UAV data with three machine learning algorithms. The aim was driven by the the fact that Philippines' rice production is still low compared to neighbouring countries and it is necessary to increase their rice production by increasing the yield on existing land. Monitoring and estimating both plant dry weight and rice height is one way to predict and estimate the rice yield. In this study, the estimations were done using three machine learning algorithms, ANN, SVM, and RF across two seasons of data from a well managed experimental set up at IRRI.

5.1. Variable selection and variable importance

This study confirms that the selection of input variables in the three different machine learning algorithms affects the result of model estimations. The selection of variables in this study was done by establishing relationships between field measured and UAV spectral data. The first relationship was between total plant dry weight and spectral reflectance, to determine which bands had the highest correlation with plant dry weight. The result showed that the near infra-red band has the highest correlation with total plant dry weight, followed by the red-edge band. Hence, these two bands were the first to be selected as input variables for the machine learning algorithms. The second relationship was between total plant dry weight and vegetation indices. There were six vegetation indices generated from multispectral bands calculations. MTVI2 was the vegetation index with the highest correlation with the total plant dry weight, therefore it was chosen as an input variable for the model.

After running the models with different sets of input variables, resulting in different model performance, using variable importance function (`varImpPlot`) in R, to see which variable had the highest degree of importance, or the variable which most affect the model's performance. Table 15 shows the ranking of variable importance with MTVI2 on 14 June 2016 (DAT 42) as the most important variable of the model. The second until the fifth most performed variables are red edge on 5 July 2016 (DAT 62), near infrared on 5 July 2016 (DAT 62), near infrared on 3 August 2016 (end of the season), and near infrared on 14 June 2016 (DAT 42). The majority of the important variables list are filled with MTVI2, average reflectance of red edge, and near infrared bands. This finding is in line with a study by (Osborne, Schepers, Francis, and Schlemmer (2002) mentioned that near infrared band reflectance were used to estimated biomass without water stress.

5.2. Performance of ANN, SVM regression, and RF algorithms in dry weight estimation

Answering the first research question about which machine learning algorithm performs the best in estimating the dry weight for 2016 EWS and 2017 DS data, resulted in different algorithms for different seasons. The results for 2016 EWS shows that the SVM regression algorithm performed the best with $R^2 = 0.75$, RMSE = 639 kg/ha, and NRMSE = 10.0% with Model 3 as the input variables. This finding with SVM method overperformed RF is rarely to be found in other studies in estimating dry weight in rice. A study comparing the SVM regression to RF algorithm in estimating dry weight in rice was done by Ndikumana et al., (2018). His findings are more in line with our finding in 2017 DS data that showed RF was the best algorithms in 3 out of 4 models in comparison to SVM and ANN algorithms. Model 3 produced the best accuracies with $R^2 = 0.88$, RMSE = 671 kg/ha, and NRMSE = 9.06%. Meanwhile Ndikumana et al., (2018) obtained model accuracy of $R^2 = 0.9$ with RMSE = 1620 kg/ha in estimating rice dry weight.

However, in our models, the SVM and RF algorithms performed similarly in the 2017 DS. This also happened in the 2016 EWS. These results suggest that there is no distinct superiority of one machine learning algorithm over another. Data skewness and model overfitting might be the reasons of the slightly different and better performance between machine learning algorithms (Horning, 2010). Overfitting is when our model does better on the training set than on the testing set, which is not the case in our models. Table 12 and Table 14 show the summary of training or calibration result compare to validation or testing result in 2016 EWS and 2017 DS.

Moreover, variation in weather condition from season to season might be one of the reasons that affect the models performance. Yang, Peng, Laza, Visperas, and Dionisio-Sese (2008) studied the rice yield gap between dry and wet season grown in LTCCE, IRRI. The finding was that higher grain yield was achieved in dry season compare to wet season, meaning greater dry weight accumulation. They discussed how high solar radiation due to less precipitation and lower atmospheric temperature in dry season are affected the yield performance, in which yield performance is determined by dry weight production and harvest index.

5.3. Negative findings of height data

This study finds that both UAV derived products for plant height, the DSM and point cloud metrics have low correlations to the field measured height data. Looking into the analysis in Section 4.4 Figure 21, the field measured height has high variability within the dataset, compare to DSM height data which has very low variability. The reason behind this is how the field height data was measured. The height was measured four times with four different plants in one field plot by stretching the tallest leaf. Then averaging the four height measurements to represent one height for one field plot. First, by stretching and measuring the tallest leaf (called the flag leaf) within one plant itself is not a good representation of the crop height (if we consider plant height to be a component of canopy height). Comparing to DSM data which generated from UAV images from a particular height above the ground and capturing the surface of the field, it is generalizing the height data, but it is representing the whole plot canopy. Therefore, it is logical that DSM data does not have a high correlation with the field measured data since they are representing two different things.

The analysis with DSM data then brought us to analyse the point cloud data by deriving height metrics. A study by Moeckel et al. (2018) found that the advantages of using multiple height metrics in estimating height is that it provides a more complete range of height information. Although several metrics are highly correlated with one another, this can be handled by the machine learning algorithms since machine learning can handle intercorrelated variables. There are 12 metrics derived from the point cloud and the correlations between these metrics to field measured data were also very poor, as shown in Section 4.5. Based on these findings, it was decided to not pursue further the estimation of heights using DSM or point cloud data.

However, since the height and dry weight field measured data shows high correlations (Section 4.1.1), then an experiment was done by adding height metrics into the dry weight model. This experiment was to understand whether the addition of height metrics will improve the dry weight model performance. The result showed the model had a worse performance after adding the height metrics. The findings in this study using the height data or relating to height data are not positively answering the second research question.

One concerning point of the field measurement of rice height is the small number of measurements, which is only conducted once at a few days before the harvest day. Given the variation in the field measured heights, it may be the case tht four samples per plot is not enough to represent the rice crop height. Also, during the end of the rice growing season, the rice plants tend to bend over with the weight of the full rice panicles. Therefore, the field measurement which stretch the tallest leaf is uncomparable with DSM or point cloud data, which is more representative of the rice crop canopy height.

Another challenge that needs to be addressed is the method of measuring ground truth of plant height. The field measurement of rice height that was done in this study, by taking four samples within subplot and average it, was not representing the whole subplot. A protocol of ground truth measurement of crop height needs to be developed. Our recommendation is to include an additional field measure of the rice canopy height at harvest time if this is deemed to be a useful parameter for rice yield estimation. This could then be compared to the DSM or point cloud metrics

Alternative ways to remotely sense the height of the rice crop can also be considered. One way of indirect measurement is by using a ground terrestrial laser scanner (TLS). As stated by Lumme et al. (2008), TLS is a promising method of estimating dry weight and crop height, especially small crops like barley, oat, and wheat. The precision of the beam is up to level of a few millimetres. Tilly et al. (2014) measured rice height and dry weight estimation using TLS, reporting an R^2 of 0.91 between TLS-derived height and manually measured rice height. A study by Bareth et al. (2016) compared UAV- and TLS- derived plant height for agricultural crop monitoring based on polygon grids and showed that the correlation between the two measurements was high with an R^2 of 0.91. However, the correlation is lower during the later growth stages. Their assumption was this happened due to different viewing geometries, while UAV is a nadir view and TLS provides an oblique view. These different viewing angles result in high variance of UAV-derived plant height and also lower mean of plant height values.

5.4. Limitations

The followings are limitations that we found in this study that restrict the extent of the findings.

- Limitations of the input variables: if we include all the input variables, from the historical plant dry weight, the average of multispectral bands reflectance, and all the vegetation indices, there are 124 variables in total. After the correlation analysis between plant dry weight and spectral data and vegetation indices, there are many more set combinations of input variables other than the four models developed in this study. Other vegetation indices from our list can be an input to the model.
- Limitations of field measured height and UAV-derived height data: due to the low correlations between field measured height data and the DSM data or height metrics derived from point cloud, the estimation of rice height was not pursue further. The high variability within the field measured height was not represent.
- Limitations of machine learning model: the models developed in this study were all using random parameters that the SVM and RF algorithms selected the best for models, which produced lowest the RMSE. Except ANN that we defined the number of hidden layer. Therefore, the results were more driven by the model itself, not ourselves that experience tuning in the parameters.

5.5. Recommendation

From our experience in this study, the followings are some recomendations that can be considered in the future work.

- To explore more vegetation indices that have been identified in previous studies as being suitable in estimating dry weight and height. This would require UAV data with more bands or different bands to the four used here.
- To develop height estimation model even though the low correlations between field measured height and UAV-derived height data. This would require a different protocol for measuring rice crop height that is comparable to the canopy height retrieved from the DSM or point cloud.
- To explore and experiment the parameters of machine learning algorithms and find which set of parameters works the best. This would require further exploration of the parameter space and the model variation in model performance.

6. CONCLUSIONS

The main objective of this study was to accurately estimate the dry weight and height of the rice crop using UAV data obtained during the wet season 2016 (2016 EWS) and dry season 2017 (2017 DS) in the IRRI Experimental Station, Philippines. Several conclusions to answer two research questions are presented as below.

1. Which machine learning algorithm has the best accuracy (in terms of R^2 and RMSE) for estimation of dry weight in the wet and dry season?

Machine learning algorithms that produce the best accuracy in terms of R^2 and RMSE are different for 2016 EWS and 2017 DS. For the 2016 EWS dataset with 72 samples, SVM regression produced the best model with $R^2 = 0.75$ and RMSE = 639 kg/ha (NRMSE = 10.4%). As for the 2017 DS dataset with 70 samples, the RF method produced the best accuracy in estimating the dry weight with R^2 of 0.88 and RMSE of 671 kg/ha (NRMSE = 9.1%). There was no single best algorithm in estimating dry weight across seasons.

2. Which machine learning algorithm has the best accuracy (in terms of R^2 and RMSE) for estimation of rice height in the wet and dry season?

Due to the low correlations between DSM height values and the height field measured data, also the low correlations between the height metrics and the height field measured data, the assessment of estimating rice height using machine learning was dropped from this study. An additional study was done by adding the height metrics into the dry weight model to see whether it would improve the model's accuracy. The result shows a decrease in dry weight model's accuracy, with R^2 dropped to 0.57 and RMSE = 867 kg/ha. The negative finding is attributed to the field measured height representing a rice crop height (based on stretching up the tallest leaf) that could never be observed with remote sensing methods.

To conclude, despite some limitations, this study has estimated the dry weight of rice from IRRI LTCCE rice field using UAV data from 2016 EWS and 2017 DS data and comparing three machine learning algorithms to produce a model with the best accuracy. The UAV data was found to be an adequate dataset in deriving important crop parameters information throughout the rice growth period. We tried to estimate the rice height but failed to do because of very low correlations between field measured height and UAV-derived height data. Some recommendations have been proposed and can be considered for the further studies.

LIST OF REFERENCES

- Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3254–3264.
<https://doi.org/10.1109/JSTARS.2016.2561618>
- Bareth, G., Bendig, J., Tilly, N., Hoffmeister, D., Aasen, H., & Bolten, A. (2016). A comparison of UAV- and TLS-derived plant height for crop monitoring: Using polygon grids for the analysis of crop surface models (CSMs). *Photogrammetrie, Fernerkundung, Geoinformation*, 2016(2), 85–94.
<https://doi.org/10.1127/pfg/2016/0289>
- Bendig, J. V. (2015). Unmanned aerial vehicles (UAVs) for multi-temporal crop surface modelling (University of Cologne). Retrieved from http://kups.ub.uni-koeln.de/6018/1/Bendig_PhD_2014_Ort&Datum_final_noCV.pdf
- Bonfil, D. J. (2017). Wheat phenomics in the field by RapidScan: NDVI vs. NDRE. *Israel Journal of Plant Sciences*, 64, 3–4.
- Breiman, L. (2001). Random Forests. In *Machine Learning* (pp. 5–32).
<https://doi.org/10.1201/9780367816377-11>
- Bryll, R., Gutierrez-Osuna, R., & Quek, F. (2003). Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 36(6), 1291–1302.
[https://doi.org/10.1016/S0031-3203\(02\)00121-8](https://doi.org/10.1016/S0031-3203(02)00121-8)
- Castro, W., Oblitas, J., Santa-Cruz, R., & Avila-George, H. (2017). Multilayer perceptron architecture optimization using parallel computing techniques. *PLoS ONE*, 12(12), 1–17.
<https://doi.org/10.1371/journal.pone.0189369>
- Curran, P. (1980). Multispectral remote sensing of vegetation amount. *Progress in Physical Geography*.
<https://doi.org/10.1177/030913338000400301>
- Devia, C. A., Rojas, J. P., Petro, E., Martinez, C., Mondragon, I. F., Patino, D., ... Colorado, J. (2019). High-Throughput Biomass Estimation in Rice Crops Using UAV Multispectral Imagery. *Journal of Intelligent and Robotic Systems: Theory and Applications*. <https://doi.org/10.1007/s10846-019-01001-5>
- Fritsch, S., Guenther, F., Wright, M. N., Suling, M., & Mueller, S. M. (2019). Package 'neuralnet.'
- Gitelson, A., & Merzlyak, M. N. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology*, 143(3), 286–292.
[https://doi.org/10.1016/S0176-1617\(11\)81633-0](https://doi.org/10.1016/S0176-1617(11)81633-0)
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies : Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90, 337–352.
<https://doi.org/10.1016/j.rse.2003.12.013>
- Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., ... Yang, X. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, 15(1), 1–19.
<https://doi.org/10.1186/s13007-019-0394-z>
- Horning, N. (2010). Random Forests: An algorithm for image classification and generation of continuous fields data sets. *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences 2010*, 1–6. <https://doi.org/10.5244/C.22.54>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*.
<https://doi.org/10.1177/02632760022050997>
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25, 295–309.
- IRRI. (2019). *LTCCE Summarized Final Data 2016EWS and 2017DS*. Manila.
- Jin, Xiu liang, Diao, W. ying, Xiao, C. hua, Wang, F. yong, Chen, B., Wang, K. ru, & Li, S. kun. (2013). Estimation of Wheat Agronomic Parameters using New Spectral Indices. *PLoS ONE*, 8(8).
<https://doi.org/10.1371/journal.pone.0072736>
- Jin, Xiuliang, Yang, G., Xu, X., Yang, H., Feng, H., Li, Z., ... Lan, Y. (2015). Combined Multi-Temporal Optical and Radar Parameters for Estimating LAI and Biomass in Winter Wheat Using HJ and RADARSAR-2 Data. *Remote Sensing*, 7(May 2014), 13251–13272.
<https://doi.org/10.3390/rs71013251>
- Jin, Y. Q., & Liu, C. (1997). Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks. *International Journal of Remote Sensing*, 18(4), 971–979.

- <https://doi.org/10.1080/014311697218863>
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecological Society of America*, 50(4), 663–666.
- Karimi, Y., Prasher, S. O., Madani, A., & Kim, S. (2008). Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. *Canadian Biosystems Engineering / Le Genie Des Biosystems Au Canada*, 50.
- Khun, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Hunt, T. (2020). *Package ‘caret’ R topics documented*.
- Lumme, J., Karjalainen, M., Kaartinen, H., Kukko, A., Hyypä, J., Hyypä, H., ... Kleemola, J. (2008). Terrestrial laser scanning of agricultural crops. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 37.
- Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, 29(3), 617–663.
<https://doi.org/10.1080/01431160701352154>
- Moeckel, T., Dayananda, S., Nidamanuri, R. R., Nautiyal, S., Hanumaiah, N., Buerkert, A., & Wachendorf, M. (2018). Estimation of vegetable crop parameter by multi-temporal UAV-borne images. *Remote Sensing*, 10(5), 1–18. <https://doi.org/10.3390/rs10050805>
- Murty, M. N., & Raghava, R. (2016). *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*. Sw: Springer.
- Ndikumana, E., Minh, D. H. T., Nguyen, H. T. D., Baghdadi, N., Courault, D., Hossard, L., & Moussawi, I. El. (2018). Estimation of rice height and biomass using multitemporal SAR Sentinel-1 for Camargue, Southern France. *Remote Sensing*, 10(9), 1–18. <https://doi.org/10.3390/rs10091394>
- Osborne, S. L., Schepers, J. S., Francis, D. D., & Schlemmer, M. R. (2002). Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and water-stressed corn. *Crop Science*, 42(1), 165–171. <https://doi.org/10.2135/cropsci2002.0165>
- Ponce, E. R., & Inocencio, A. B. (2017). *Toward a More Resilient and Competitive Philippine Rice Industry: Lessons from the Past Three Decades*. Retrieved from <http://books.irri.org/2017-March-Rice-Report-Lessons-from-the-past-3-decades.pdf>
- Prabhakara, K., Hively, W. D., & Mccarty, G. W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *International Journal of Applied Earth Observations and Geoinformation*, 39, 88–102.
<https://doi.org/10.1016/j.jag.2015.03.002>
- Reisi-Gahrouei, O., Homayouni, S., McNairn, H., Hosseini, M., & Safari, A. (2019). Crop biomass estimation using multi regression analysis and neural networks from multitemporal L-band polarimetric synthetic aperture radar data. *International Journal of Remote Sensing*, 40(17), 6822–6840.
<https://doi.org/10.1080/01431161.2019.1594436>
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. *NASA Special Publication 351*, 309.
- senseFly. (2014). *User Manual multiSPEC 4C camera Revision 1 / August, 2014 Copyright © 2010-2014 senseFly Ltd*. Lausanne: senseFly Ltd.
- Silleos, N. G., Alexandridis, T. K., Gitas, I. Z., Silleos, N. G., & Alexandridis, T. K. (2008). *Vegetation Indices: Advances Made in Biomass Estimation and Vegetation Monitoring in the Last 30 Years*. 6049.
<https://doi.org/10.1080/10106040608542399>
- Skidmore, A. K., Turner, B. J., Brinkhof, W., & Knowles, E. (1997). Performance of a neural network: Mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 63(5), 501–514.
- Sonika, & Rathi, P. (2018). Biomass estimation at ICESat/GLAS footprints using support vector regression algorithm for optimization of parameters. *Advances in Intelligent Systems and Computing*, 584, 101–111. https://doi.org/10.1007/978-981-10-5699-4_11
- Tilly, N., Hoffmeister, D., Cao, Q., Huang, S., Lenz-Wiedemann, V., Miao, Y., & Bareth, G. (2014). Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice. *Journal of Applied Remote Sensing*, 8(1), 083671.
<https://doi.org/10.1117/1.jrs.8.083671>
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory* (Vol. 3). Retrieved from <http://repositorio.unan.edu.ni/2986/1/5624.pdf>
- Viljanen, N., Honkavaara, E., Näsi, R., Hakala, T., Niemeläinen, O., & Kaivosoja, J. (2018). A novel

machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture*, 8(5).

<https://doi.org/10.3390/agriculture8050070>

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop Journal*, 4(3), 212–219.

<https://doi.org/10.1016/j.cj.2016.01.008>

Yang, W., Peng, S., Laza, R. C., Visperas, R. M., & Dionisio-Sese, M. L. (2008). Yield gap analysis between dry and wet season rice crop grown under high-yielding management conditions. *Agronomy Journal*, 100(5), 1390–1395. <https://doi.org/10.2134/agronj2007.0356>