

Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning

Vasit Sagan^{a,b,*}, Maitiniyazi Maimaitijiang^{a,b}, Sourav Bhadra^{a,b}, Matthew Maimaitiyiming^{a,c}, Davis R. Brown^{a,d}, Paheding Sidike^e, Felix B. Fritsch^f

^a Geospatial Institute, Saint Louis University, St. Louis, MO 63108, USA

^b Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO 63108, USA

^c Division of Food Sciences, University of Missouri, Columbia, MO 65211, USA

^d Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^e Department of Applied Computing, Michigan Technological University, Houghton, MI 49931, USA

^f Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA



ARTICLE INFO

Keywords:

PlanetScope
WorldView-3
Deep learning
Convolutional neural network
ResNet
Artificial intelligence
Food security

ABSTRACT

Agricultural management at field-scale is critical for improving yield to address global food security, as providing enough food for the world's growing population has become a wicked problem for both scientists and policy-makers. County- or regional-scale data do not provide meaningful information to farmers who are interested in field-scale yield forecasting for effective and timely field management. No studies directly utilized raw satellite imagery for field-scale yield prediction using deep learning. The objectives of this paper were twofold: (1) to develop a raw imagery-based deep learning approach for field-scale yield prediction, (2) investigate the contribution of in-season multitemporal imagery for grain yield prediction with hand-crafted features and WorldView-3 (WV) and PlanetScope (PS) imagery as the direct input, respectively. Four WV-3 and 25 PS imagery collected during the growing season of soybean were utilized. Both 2-dimensional (2D) and 3-dimensional (3D) convolution neural network (CNN) architectures were developed that integrated spectral, spatial, temporal information contained in the satellite data. For comparison, hundreds of carefully selected spectral, spatial, textural, and temporal features that are optimal for crop growth monitoring were extracted and fed into the same deep learning model. Our results demonstrated that (1) deep learning was able to predict yield directly using raw satellite imagery to the extent that was comparable to feature-fed deep learning approaches; (2) both 2D and 3D CNN models were able to explain nearly 90% variance in field-scale yield; (3) limited number of WV-3 outperformed multi-temporal PS data collected during entire growing season mainly attributed to RedEdge and SWIR bands available with WV-3; and (4) 3D CNN increased the prediction power of PS data compared to 2D CNN due to its ability to digest temporal features extracted from PS data.

1. Introduction

With a changing climate, agricultural management is becoming increasingly important to provide enough food for the world's growing population. To meet this demand, improving crop yield, a measure of grains or seeds generated per unit of area, sustainably continues to be of great interest to both farmers and policymakers (Pantazi et al. 2016). Even though there are areas for cropland expansion, emphasis should be placed on preventing yield loss and increasing output within the current agricultural footprint to reduce the detrimental effects of agricultural expansion on the environment. Near real-time monitoring of crop

growth and vigor at individual field-scale is expected to benefit this effort (Maimaitijiang et al. 2020a), as preventive measures (fertilization and irrigation) against yield loss can be taken to boost agricultural productivity when and where most needed. There is an urgent need to develop cost-effective, artificial intelligence powered approaches that can provide near-real-time, intra-field phenology, and crop condition to increase both yield and resource use efficiency.

Remote sensing provides non-destructive, fast, and economical monitoring of the earth's surface over a large area, and are recognized as a valuable tool for yield prediction (Becker-Reshef et al. 2010). Numerous yield prediction studies that employed a single or

* Corresponding author at: Geospatial Institute, Saint Louis University, St. Louis, MO 63108, USA.

E-mail address: vasit.sagan@slu.edu (V. Sagan).

combination of multiple features (spatial, temporal and spectral) extracted from satellite imagery have been reported since the 1970s (Battude et al. 2016; Johnson 2014; Mariotto et al. 2013). The methods in these studies can be summarized into two broad categories: crop growth modeling and machine learning.

Crop growth models (Batchelor et al. 2002) assimilated with remote sensing data have shown promise to improve yield prediction (Doraiswamy et al. 2004). The assimilation process involves the coupling of crop growth models that are site-specific with remotely sensed soil moisture and leaf area index (LAI) (Betbeder et al. 2016; Wang et al. 2011). For example, Ines et al. (2013) demonstrated that the Decision Support System for Agro-technology Transfer – Cropping System Model (DSSAT-CSM)-Maize integrated with LAI and soil moisture from remote sensing observations improved yield prediction with increased correlation and minimized errors compared to the results of DSSAT-CSM modeling only. However, the performance of such assimilation is limited by scale mismatch between coarse spatial resolution satellite images and crop models that operate at field scale (Huang et al. 2019). To overcome this limitation, World Food Studies Simulation Model (WFOST) was integrated with radiative transfer simulation models (i.e., PROSAIL) (Huang et al. 2015) or hydrology-crop growth model (Li et al. 2014c) providing a modest improvement over previous studies. However, daily meteorological data are the main input for crop growth simulations that are not always readily available at optimal spatial and temporal scales.

Machine learning is a set of statistical methods designed for solving specific tasks (e.g., classification or regression) by automatically detecting patterns and anomalies from data and performing decision-making. Linear regression models are based on the empirical relationship between crop yield and abiotic factors that dictate growth including precipitation, temperature, and soil hydraulic conditions (Kern et al. 2018; Qian et al. 2009). This approach often employs multivariate regression models (Becker-Reshef et al. 2010; Tucker et al. 1980). Spectral indices derived from satellite data, i.e. Normalized Difference Vegetation Index (NDVI), when combined with those factors, were found to improve prediction of maize (*Zea mays*) and soybean (*Glycine max*) yields (Al-Gaadi et al. 2016; Prasad et al. 2006). A variety of spectral indices from spaceborne optical sensors such as NOAA AVHRR (Quarmby et al. 1993), SPOT-VEGETATION (Kowalik et al. 2014), Landsat and MODIS (Doraiswamy et al. 2004; Ren et al. 2008) have been applied for yield prediction. However, plant growth and grain production processes are complex and nonlinear in nature, which is affected by field management and crop and environmental interactions. Thus, they cannot be effectively modeled by linear statistical regression approaches.

Due to its ability to capture both linear and nonlinear characteristics of agricultural systems with the growing availability of data, machine learning has significantly gained the trustworthiness of remote sensing-based yield estimation (Cai et al. 2019; Maimaitijiang et al. 2020b; Maimaitiyiming et al. 2019). Examples include partial least squares regression (Rischbeck et al. 2016), Random Forest Regression (RFR) (Aghighi et al. 2018), and Support Vector Regression (SVR) (Gomez et al. 2019; Khanal et al. 2018). A noteworthy development in nonlinear statistical learning is that models (e.g., RFR) are becoming more generalized and available in cloud platforms (e.g., Google Earth Engine), allowing continental-scale mapping of crop type and yield using multi-modal satellite data available on the cloud (Burke and Lobell 2017; Jin et al. 2019). On the other hand, a suite of artificial neural networks, such as Bayesian Neural Networks (BNN) (van Hinsbergen et al. 2009) and multi-layer feed-forward neural networks (Gevrey et al. 2006) have

been used to forecast crop yields (Fieuza et al. 2017; Johnson et al. 2016). These approaches and models mostly have been developed using spectral, spatial, and/or texture information derived from remote sensing data, independent of crop growth modeling.

Deep Learning, a subfield of machine learning, has become a popular choice of methods for yield prediction in recent years (Jiang et al. 2019; Ma et al. 2017; Nevavuori et al. 2019) due to its superior performance over conventional methods (Sidike et al. 2019). This is attributed partly to the latest developments in algorithms and computing resources, and the fact that deep learning is able to automatically learn and abstract both linear and nonlinear hierarchical features from the data using a multi-layer architecture. By using spectral and spatial features (Vergara-Diaz et al. 2016) derived from remote sensing data or soil and meteorological data (Khaki and Wang 2019), previous studies demonstrated that deep learning has significant potential for predicting crop yield. Jiang et al. (2019) implemented a long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) model to predict county-level corn yields using features such as crop phenology, meteorology, and wide dynamic range vegetation index (WDRVI) from remote sensing data. Most recently, a multi-modal deep learning model that integrates spectral, structure, thermal, and texture (i.e., GLCM: gray-level co-occurrence matrix) information from Unmanned Aerial Vehicle (UAV) data was developed for soybean yield prediction (Maimaitijiang et al. 2020b). These studies utilized hand-crafted features, namely, spectral indices, and spatial and textural features that were developed to be optimal to monitor crop growth. Developing hand-crafted features requires a priori knowledge of remote sensing data as well as light and landscape interactions. Features representing ideal variables for yield prediction are selected based on correlation analysis between the features and yield through end-to-end deep learning frameworks. Often there are hundreds or thousands of features to prepare, analyze, and select for modeling, significantly limiting the transferability of machine learning models over space and time to address global food security challenges in an era of big data.

To the best of our knowledge, there have been two studies on crop yield forecasting in which satellite imagery was used as direct input to a deep learning model. You et al. (2017) introduced a dimensionality reduction technique that allows training of a CNN network to automatically learn spatio-temporal structure of the data and extract useful features. Using a Gaussian process modeling, LSTM network and CNN, 12-year MODIS data were converted to histograms of intensity, time, and depth, which were then used to predict county-scale soybean yield data. However, this still requires abstracting and extracting some features (e.g., spatial, textural) to histograms at the cost of spatial information which is critical for yield prediction (Yang et al. 2019). Russello (2018) extended You et al. (2017)'s work to three dimensional CNN (3D CNN) leveraging spatial and temporal features for soybean yield prediction and found a modest improvement in root mean square error (RMSE). Nevertheless, it is still unclear if the generalization or transferability of deep learning methods is robust enough for yield estimation at field-scale, especially 3D CNN algorithms which have shown great potential in various applications due to its ability to learn from spatio-temporal data. Additionally, county-scale data do not provide meaningful information to farmers who are interested in field-scale yield prediction for effective and timely field management. To date, deep learning-based yield prediction using remote sensing imagery as direct input is less explored, particularly, to the best of our knowledge, no studies have directly utilized raw satellite imagery for field-scale yield prediction using deep learning nor compared it with feature-based yield prediction, this also likely due to the lacking of

fine scale yield data.

Availability of very high-resolution (VHR) satellite sensors, i.e., WorldView-3 (WV) and PlanetScope (PS), have tremendous potential for estimating yield at field-scale. WV has very high spatial resolution at panchromatic (0.31 m) and VNIR (1.24 m) bands, and SWIR bands are available at 7.5 m spatial resolution (Sidiike et al. 2019). Although PS has fewer bands and lower spatial resolution (four bands in VNIR spectrum at 3 m spatial resolution) compared to WV, it may be advantageous for crop monitoring due to its frequent revisit time (less than one day). Crop phenological stages (e.g., seed-filling stage (R5) for soybean) was found to be critical for predicting yield (Ma et al. 2001). In addition, remote sensing data with red edge and near-infrared (NIR) band (800 nm) were found to be most effective for predicting yield (Zhou et al. 2017). One may expect that PS provides a unique opportunity to predict terminal yield by capturing key development stages with its daily revisit frequency. However, no studies investigated the pros and cons of WV data with limited temporal coverage versus frequent (daily) observations of PS data for field-scale yield prediction.

The objectives of this study are to (1) propose a raw imagery-based end-to-end deep learning approach for field-scale yield prediction, (2) investigate the contribution of in-season multitemporal imagery for grain yield prediction with hand-crafted features and raw input imagery, respectively.

2. Test site and data

2.1. Experimental design

Soybean and maize experimental fields were established at the

University of Missouri Bradford Research Center (BRC) near Columbia, Missouri, USA in 2017 (Fig. 1). The study site has a humid continental climate with average monthly growing season temperatures that peaked at 25.0 °C in July and were lowest at 13.8 °C in October (Maimaitijiang et al. 2020b). Prior to soybean planting, the field was disked to approximately 0.15 m depth followed by one pass with a cultivator. Soybean was planted on 19 May 2017 in 0.76-m wide rows at a density of 400,000 seeds ha⁻¹. The field (approximately 61 m wide and 220 m long) was planted with ‘Pana’, ‘Dwight’, ‘AG3432’ cultivars. The design encompassed four replications, with each consisting of an irrigated and a rainfed main plot with the three cultivars randomly arranged within each main plot. The cultivar subplots were either 17 m × 21 m or 21 m × 27 m in size. No fertilizer was applied and weeds were controlled with pre- and post-emergence herbicide applications.

Corn hybrids were planted on 26 May 2017 in a lattice design with three replications in a rainfed (Corn-1) and an irrigated (Corn-2) field (Fig. 1). A total of 59 and 37 modern corn hybrids were grown in rainfed and irrigated fields, respectively. They are referred to as hybrids in this manuscript due to the proprietary nature of the genotypes. Hybrids were planted in plots that were four rows wide (0.76 m row distance) and 8.2 m long at a density of 74,130 kernels ha⁻¹ in the rainfed and 93,900 kernels ha⁻¹ in the irrigated field. The rainfed field was fertilized with 180 kg N ha⁻¹ whereas the irrigated field was fertilized with 269 kg N ha⁻¹. Weeds were controlled using pre- and post-emergence herbicide applications. The center two rows of each plot were harvested with a small plot combine on 3 October 2017 (irrigated) and 9 October 2017 (rainfed).

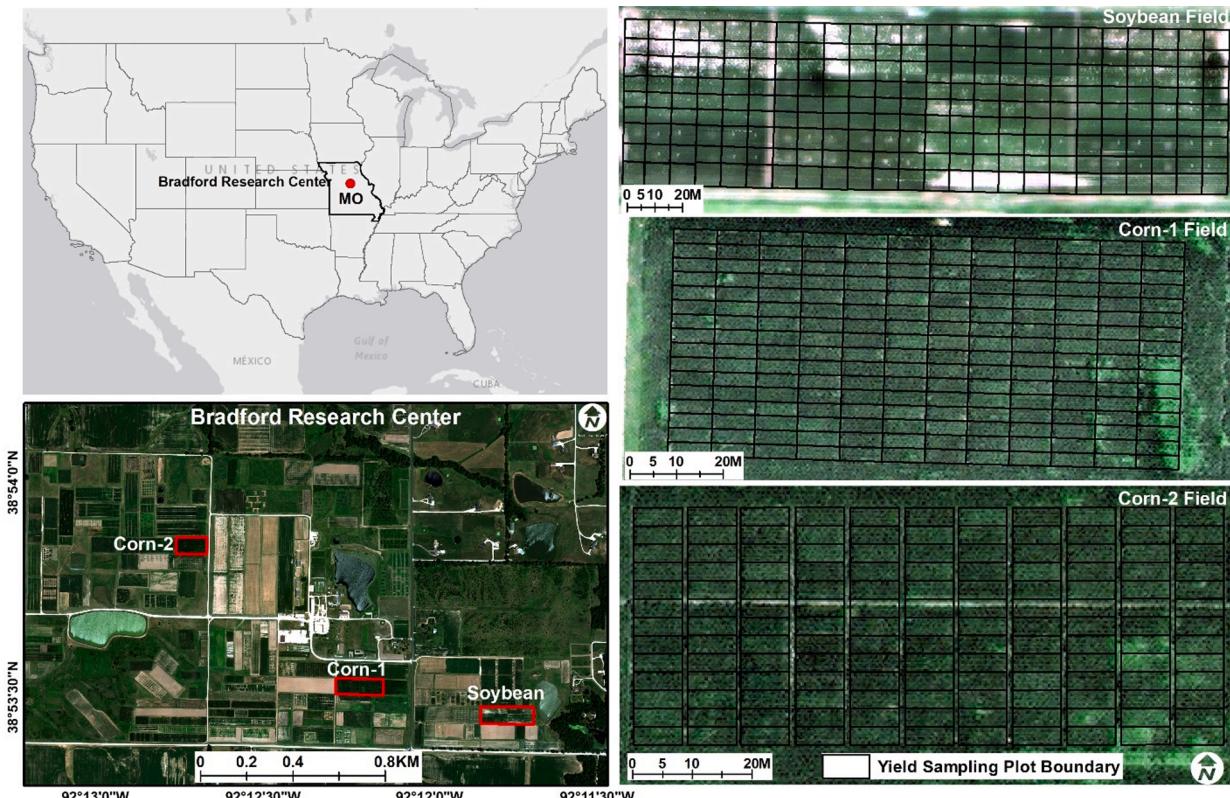


Fig. 1. Test site location and field layout. Solid black lines on each field indicate the plot boundary.

2.2. Data acquisition

2.2.1. Field data collection

A small-plot combine (ALMACO SPC40, ALMACO, Nevada, IA) was employed for harvesting soybean in October 2017. Two-row subplots of either 8.1 m or 9.1 m length were harvested and weighed, which results in a total of 960 12.3 m² or 13.8 m² yield plots within the entire field. To match the pixel size of PS and WV satellite imagery, 4 or 5 subplots were combined to make one plot, which resulted in 216 yield plots that were used as the total data samples for soybean yield prediction (Fig. 1). The soybean grain yield from each plot (kg ha⁻¹) was normalized to a moisture content of 13.0%.

For corn, the center two rows of each plot were harvested with a small-plot combine on 3 October and 9 October 2017, in the irrigated and the rainfed fields, respectively. There were a total of 293 plots with valid yield data for the cornfields, 176 plots in the rainfed, and 117 plots in the irrigated field, respectively. The corn grain yield from each plot in kg ha⁻¹ was normalized to a moisture content of 15.5%.

2.2.2. WorldView-3 imagery and pre-processing

Four sets of cloud-free WorldView-3 (WV) (Digital Globe Inc., Westminster, Colorado, USA) imagery were tasked between 16 July and 04 September 2017 (Table 1) and were used in this study. The WV images contain a 0.3 m spatial resolution panchromatic band, eight 1.2 m spatial resolution Visible Near Infrared (VNIR), and eight 7.5 m resolution Short Wave Infrared (SWIR) bands. The images were atmospherically corrected to derive surface reflectance (Sagan et al. 2019; Sidike et al. 2019) using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm (Guide, 2009) with ENVI 5.5.3 software (L3Harris Geospatial, Boulder, Colorado, USA). WV-3 data, including VNIR and SWIR, were pan-sharpened with Gramm-Schmidt Pan Sharpening tool of ENVI software (Hartling et al. 2019) to 0.3 m resolution using the panchromatic band. Multitemporal co-registered WV-3 images were layer stacked to produce the final image cube for further analysis.

2.2.3. PlanetScope imagery and pre-processing

A total of 25 available cloud-free PS images collected between 21 June and 09 September 2017 were downloaded from Planet Labs (Table 1). These images were orthorectified, provided as TOA radiance (Level 3B), and analytics-ready surface reflectance by the vendor. The PS imagery has four bands including Blue, Green, Red, and Near-Infrared, with an orthorectified pixel size of 3.125 m (Houborg and McCabe 2018). All PS images were resampled to 0.3 m and layer stacked to a temporal image cube.

2.2.4. UAV imagery collection and pre-processing

A UAV (Unmanned Aerial Vehicle) data collection campaign was conducted on 20 July, 2017 (Table 1). The purpose of the UAV imagery collection was to accurately geolocate each harvest plot on satellite imagery by co-registering it with the very high-resolution UAV RGB imagery. The details about UAV data collection and processing are as follows:

A lightweight UAV Mavic Pro quadcopter (DJI, Shenzhen, China) was used for visual imagery collection. The flight altitude was 30 m and flight speed was 6 m s⁻¹ with a front overlap of 85% and side overlap of 85%. Mavic Pro quadcopter is a small light commercial platform that weighs 743 g and has about 30-min maximum flight time. The RGB camera on Mavic Pro has a 12.4 Mpixel CMOS sensor and 5 mm focal

Table 1

Remote sensing imagery and acquisition date.

Platforms	Spectral bands	Original Spatial Resolution	No. of Imagery	Imagery Acquisition Date (Year: 2017)
Satellite (WorldView-3)	Panchromatic (Pan) VNIR: Coast, Blue, Green, Yellow, Red, Red Edge, Near Infrared1, Near Infrared2	Pan: 0.31 m VNIR: 1.24 m SWIR: 7.5 m	4	07/16; 07/22; 08/09; 09/04
Satellite (PlanetScope)	Red, Green, Blue, NIR	3.125 m	25	06/21; 06/24; 06/25; 07/03; 07/07; 07/08; 07/09; 07/11; 07/15; 07/16; 07/17; 07/19; 07/20; 07/21; 07/22; 07/25; 07/28; 07/29; 08/04; 08/13; 08/19; 08/24; 08/25; 09/07; 09/09;
UAV (RGB)	Red, Green, Blue	0.01 m	/	07/08

length with a 78.8 field-of-view lens. The UAV RGB images were mosaicked and orthorectified using Pix4Dmapper software package (Pix4D SA, Lausanne, Switzerland) along with ground control points (GCPs) established in the field (Maimaitijiang et al. 2019). UAV orthomosaics, WV, and PS satellite data were co-reregistered to a common spatial resolution with WGS 1984 datum UTM Zone 15 N.

3. Methodology

Fig. 2 shows the overall technical workflow of this study, which starts with radiometric calibration, pan-sharpening of WV images, followed by co-registration of both WV and PS data over each plot. Hand-crafted features including various spectral and textural indices were calculated, stacked into multitemporal image cubes, which were then clipped by plot boundary shapefile to produce individual image chips representing each plot. Several conventional machine learning algorithms were implemented along with features and raw imagery-based deep learning architectures as detailed in the following sections.

3.1. Feature extraction

Plot-level average reflectance values from each band of WV and PS imagery were used as canopy spectral features. Additionally, various vegetation indices (VIs) used for yield prediction in previous literature were taken as spectral features for yield prediction (Table 2). Moreover, the GLCM (grey level co-occurrence matrix) texture features, (Haralick et al. 1973), which includes mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (SEC) and correlation (COR), were computed using the ENVI 5.4.1 software and used for grain yield prediction (Shiu and Chuang 2019b). Readers are referred to Nichol and Sarker (2011) for detailed descriptions of these features.

Table 2

List of manually extracted features used in this study.

Platforms	Features	Formulation	References
Raw Bands	Coast (C), Blue (B), Yellow (Y), Green (G), Red (R), Red-edge (RE), Near-infrared1 (NIR1), Near-infrared2 (NIR2), SWIR1, SWIR2, SWIR3, SWIR4, SWIR5, SWIR6, SWIR7, SWIR8,	The reflectance value of each band	/
Vegetation Indices for both WorldView-3 and PlanetScope	Ratio vegetation index Normalized difference vegetation index Green normalized difference vegetation index The enhanced vegetation index Enhanced Vegetation Index (2-band) Wide dynamic range vegetation index Structure Insensitive Pigment Index Normalized Ratio Vegetation Index Visible atmospherically resistance index Transformed Vegetation Index	RVI = NIR1 / R NDVI = (NIR1 - R) / (NIR1 + R) GNDVI = (NIR1 - G) / (NIR1 + G) EVI = 2.5*((NIR1-R)/(NIR1 + 6*R - 7.5*B + 1)) EVI2 = 2.5 * (NIR - RED) / (NIR + 2.5*RED + 1) WDRVI=(a*NIR1-R)/(a*NIR1 + R) (a = 0.12) SIPI = (NIR1 - B) / (NIR1 - R) NRVI = (RVI - 1) / (RVI + 1) VARI = (G - R) / (G + R - B) TVI = sqrt [(NIR1 - R) / (NIR1 + R) + 0.5]	(Tucker, 1979) (Rouse et al. 1974) (Gitelson et al. 2003) (Huete et al. 2002) (Jiang et al. 2008) (Gitelson 2004) (Penuelas et al. 1995) (Baret and Guyot 1991) (Gitelson et al. 2002) (Deering 1975)
Vegetation Indices only for WorldView-3	Optimized soil adjusted vegetation index Modified chlorophyll absorption in reflectance index Transformed chlorophyll absorption in reflectance index MCARI/OSAVI TCARI/OSAVI Normalized Difference Water Index (NDWI)	OSAVI=(NIR1-R)/ (NIR1-R + L) (L = 0.16) MCARI = [(RE-R)- 0.2*(RE-G)] *(RE/R) TCARI = 3*[(RE-R)-0.2*(RE-G) * (RE/R)] MCARI/OSAVI TCARI/OSAVI NDWI = (NIR-SWIR _i)/(NIR + SWIR _i) (i = 1,2,3,4,5,6,7,8)	(Rondeaux et al. 1996) (Daughtry et al. 2000) (Haboudane et al. 2002) (Daughtry et al. 2000) (Haboudane et al. 2002) (Gao 1996)
Texture Features	Normalized difference red-edge Red-edge chlorophyll index Gray-level co-occurrence matrix (GLCM)	NDRE = (NIR1 - RE) / (NIR1 + RE) RECI = (NIR1 /RE) - 1 MEA, VAR, HOM, COR, DIS, ENT, SEC, COR	(Gitelson and Merzlyak 1997) (Gitelson et al. 2005) (Haralick et al. 1973)

3.2. Feature selection

The selection of sensitive features is an important issue in any machine learning application because an appropriate number of selected features can reduce both model and computational complexity while improving performance. We calculated a total of 765 features from WV (72 bands, 117 VIs, and 576 texture features), and 1047 features from PS (100 bands, 275 VIs, and 672 texture features), which required reducing the dimension of the input matrix by selecting sensitive features. Therefore, we trained a random forest regression model for each training

set and calculated the Mean Decrease Impurity (MDI) score from the trained model. MDI was proposed by Breiman (2001) to evaluate the importance of a variable by averaging the sum of weighted impurity decreases for all nodes over all trees in the random forest architecture. Many studies have adopted MDI as a potential feature importance score (Georganos et al. 2018; Hariharan et al. 2018). From the distribution of feature importance scores, the 95th percentile score was selected as the threshold, i.e., the features which yielded greater than the threshold were kept as a sensitive feature, otherwise, the feature was discarded.

3.3. Feature-based modeling

Based on hand-crafted features, commonly used machine learning methods, Partial least-square regression (PLSR) (Rischbeck et al. 2016), Support Vector Regression (SVR) (Cortes and Vapnik 1995; Kuwata and Shibasaki 2016), Random Forest Regression (Aghighi et al. 2018), and a fully-connected feed-forward Deep Neural Network (DNN) (Biganzoli et al. 1998) were employed for yield prediction. The fully-connected feedforward DNNs have been increasingly applied to a range of classification and regression problems (Cai et al. 2018; Langkvist et al. 2014) with great success.

The optimized hidden layers and activation functions for the DNN model are provided in Fig. 3. In short, the overall process includes application of the Rectified Linear Unit (ReLU) activation function with the form of $f(z) = \max(0, z)$ (LeCun et al. 2015) and reducing overfitting with dropout approach (Srivastava et al. 2014), which was followed by a batch-normalization (Ioffe and Szegedy 2015). Batch-normalization was applied between hidden layers 2 and 3, 4 and 5, as well as hidden layer 6 and output layer. Parameter tuning was employed through grid search procedure while training, i.e., using batch sizes of 16, 32, 48, and 64; learning rates of 0.001, 0.004, 0.007, 0.009, 0.01, 0.03 and 0.05; dropouts of 0.20, 0.23, 0.25, 0.27, 0.30, 0.35, 0.40 and 0.50; The epoch number for training was 400. The model which yielded the minimum training loss was used for predicting the testing dataset. The Python programming environment with TensorFlow (Abadi et al. 2016) and Keras libraries (<https://keras.io/>) were used to implement the feature-based DNN.

3.4. Imagery-DL based algorithm modeling

3.4.1. ResNet-18

A common conception about CNNs in the scientific community has been that the deeper networks usually lead to better performance over shallower ones. For instance, some of the most used deeper networks, i.e., AlexNet (Krizhevsky et al. 2012), VGG-Net (Simonyan and Zisserman 2014), and GoogleNet (Szegedy et al. 2015) had 5, 19, and 22 convolutional layers, respectively. However, it has been observed that very deep networks are difficult to train due to the “vanishing gradient problem” (He et al. 2016a; Srivastava et al. 2015). The vanishing gradient problem occurs when the gradient is backpropagated to previous layers and repeated multiplication may make the gradient infinitely smaller. Therefore, as the network goes deeper, the performance of the networks gets saturated or even starts degrading rapidly (Meng et al. 2019; Srivastava et al. 2015).

Deep residual network (or ResNet) is a residual learning framework that eases the training of deeper networks (He et al. 2016a). The deep ResNet avoids such vanishing gradient problem by incorporating identity “skip-connections” which helps the gradient flowing back to the shallower layers without vanishing. By doing so, it facilitates the training of very deep networks up to thousands of layers (He et al. 2016a). Fundamental structural elements known as “residual blocks” are the core of ResNet. Fig. 4 shows the architecture of a typical residual block, called bottleneck residual block (He et al. 2016b), which performs the following equation:

$$x_l = f_l(x_{l-1}) + x_{l-1} \quad (1)$$

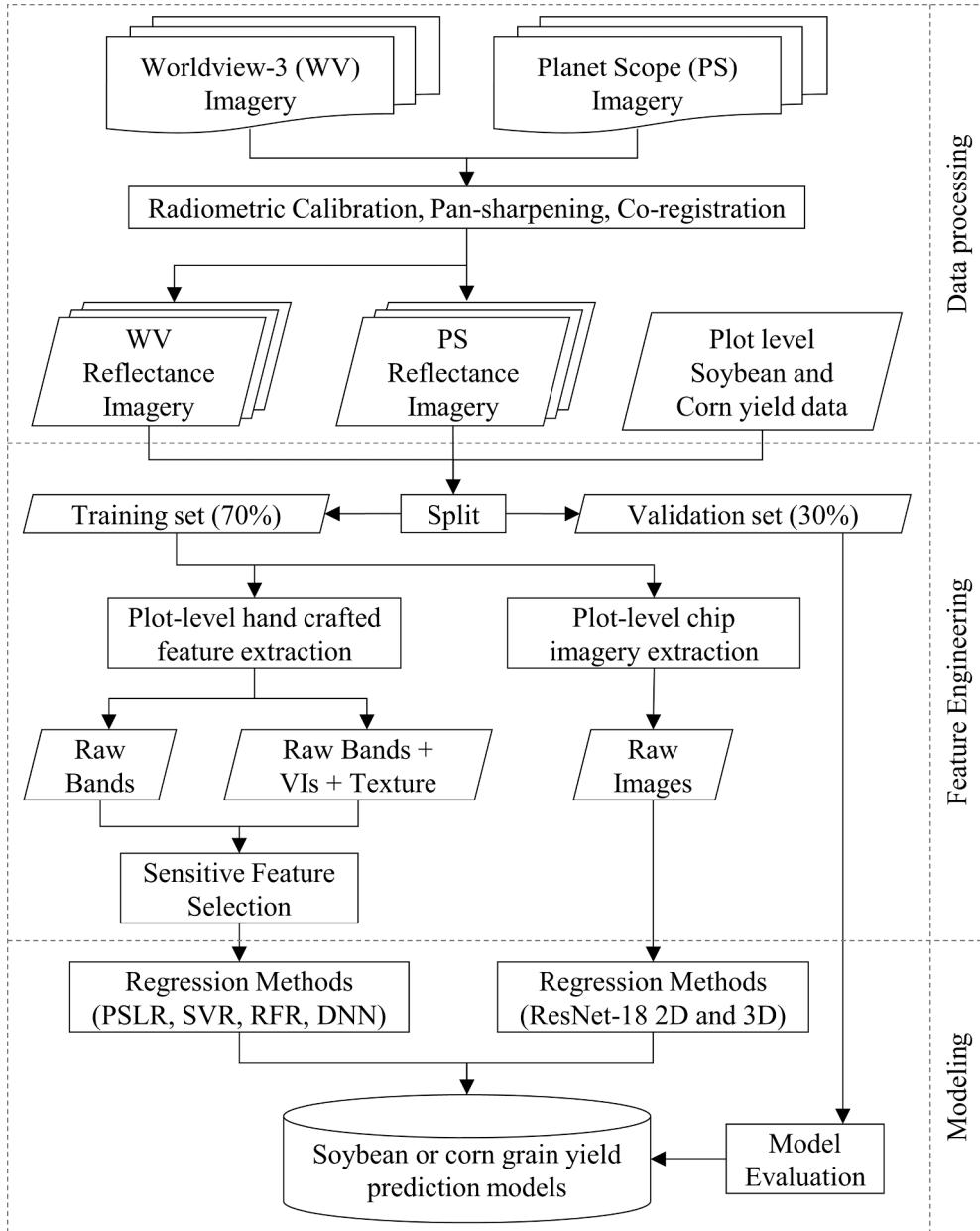


Fig. 2. Overall workflow implemented in this paper. It is worth noting that 30% of the data used as validation dataset which was not included in training the model to make sure the validation dataset was totally independent of training dataset.

where x_{l-1} and x_l are the input and output of the l th residual block, respectively and $f_l(\cdot)$ is the residual function to be learned, which consists of 3 convolutional (Conv) layers preceded by a batch normalization layer (Ioffe and Szegedy 2015) and a ReLU activation function (He et al. 2016b). The kernel size of the three Conv layers, 1×1 , 3×3 , and 1×1 are designed in a way that the first 1×1 layer reduces feature dimension and the second 1×1 layer expands it back. Hara et al. (2018) proposed a 3D version of ResNet, where the Conv layers are $1 \times 1 \times 1$, $3 \times 3 \times 3$, and $1 \times 1 \times 1$, respectively (Fig. 4b). In this study, both 2D and 3D versions of ResNet, used as ResNet 2D and ResNet 3D hereafter, have been applied, where the extra dimension of the convolution operator in the ResNet 3D is used over the temporal dimension of the raw satellite imagery.

3.4.2. Network architecture

ResNet architecture requires an input image with 224×224 spatial dimension. However, the images processed from both WV and PS had

spatial dimensions of 28×28 and thus were linearly resized into 224×224 spatial dimensions before feeding into the network. In addition, the number of bands was also different for WV and PS images since WV had 16 spectral bands for 4 days and PS had 4 spectral bands for 25 days. Table 3 summarizes the input image dimension of WV and PS in both ResNet-18 2D and 3D architectures. Fig. 5. Shows the schematic illustration of both ResNet-18 2D (He et al. 2016a) and 3D (Hara et al. 2018) architectures.

The 2D and 3D ResNet-18 architectures consist of 5 basic blocks (i.e., Conv1-5) in which each block contains 2 bottleneck blocks. The first Conv layer (Conv1) has 7^n kernel size (n is 2 for 2D and 3 for 3D) with 64 features. In ResNet 2D, Conv1 has a spatial stride of 2, whereas in ResNet 3D, it has a spatial stride of 2 and a temporal stride of 1. The output shape of Conv1 is 112×112 . The Conv1 is followed by a 3^n max-pooling layer of stride 2. The other 4 basic blocks (i.e., Conv2-5) start with 64, 128, 256, and 512 number of features with spatial stride of 2. Each basic block contains 2 bottleneck blocks where identity skip-connections are

applied. The output sizes of these basic blocks are 56×56 , 28×28 , 14×14 , and 7×7 for Conv2-5 respectively. After passing through these basic blocks, an average pooling layer and a fully connected layer with 1000 features were applied. Since yield prediction is a regression problem, a final fully connected layer with 1 feature is added by the end of the architecture. Both 2D and 3D ResNet-18 architectures were implemented using PyTorch libraries (Paszke et al. 2019).

The networks were optimized with a grid search through hyper-parameters while training, i.e., using batch sizes of 16, 32, and 64; learning rates of 0.01, 0.04, 0.001, and 0.004; momentums of 0.80, 0.85, 0.90, and 0.95; step sizes of 100, 1000, 10000, and 100000; and epochs of 400. A validation split of 30% from the training set was used to calculate the validation loss while training. The model which achieved the minimum validation loss was used for predicting the test dataset. The model training was done in a machine with Intel Xeon Platinum 8168 (2.7 GHz) 24 processors, 512 GB of RAM, and dual Nvidia RTX 8000 graphical processing unit (GPU) each with 48 GB memory.

3.5. Model evaluation

For both hand-crafted features and imagery based models, the data were split by 70% and 30% randomly. The 70% of data were used for training and the remaining 30% was only used to test the model performance. Several error metrics including the coefficients of determination (R^2), root mean square error (RMSE), and relative RMSE (RMSE%) were used as shown in the following equations.

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

$$\text{RMSE\%} = \frac{\text{RMSE}}{\bar{y}} * 100 \quad (3)$$

where y_i and \hat{y}_i refers to the measured and the predicted yield, \bar{y} is the mean of measured yield, and n is the total number of samples in the testing set.

The adaptability of the grain yield prediction model over space is crucial in dealing with spatial heterogeneity and variations due to the various environmental conditions in the field such as soil type, moisture, nutrient and so on (Maimaitijiang et al. 2020b; Rischbeck et al. 2016; Wang et al. 2020). Thus, the spatial variability of the model results over different genotypes exposed to different irrigation treatments was evaluated by using the Global Moran's I (GMI). The GMI provides an uncertainty estimation related to the spatial distribution pattern of regression residuals (Cai et al. 2014; Imran et al. 2015). The error refers to the residuals between the measured and predicted yield of each plot. Spatial autocorrelation was estimated using Global Moran's I statistic (Anselin 1995; Harries 2006), which reveals the degree of clustering of variables. The GMI values range from -1 to 1 , -1 meaning negative spatial autocorrelation, 0 is random no correlation, and $+1$ meaning positive autocorrelation. In yield prediction as it is our case, a GMI of residuals closer to 0 means the error is random and the model is robust, or vice-versa.

4. Results

4.1. Descriptive statistics of plot yields

The descriptive statistics and distribution of collected sample yields from both soybean and corn plots are shown in Fig. 6. The number of samples for corn ($n = 293$) yield was slightly higher than soybean ($n = 216$). However, the corn yield had a higher range of distribution with a higher standard deviation ($1728.66 \text{ kg/ha kg ha}^{-1}$) compared to soybean. The distribution of both yield data showed an approximately normal distribution (Fig. 6).

4.2. Yield prediction for soybean

Table 4 compares the performance of both WV and PS based models for soybean in terms of R^2 , and RMSE%. The best performance was achieved by using ResNet 2D and 3D with WV raw imagery as input ($R^2 > 0.8$ and RMSE% < 10). From PS raw imagery-based modeling, both ResNet 2D and 3D performed better than other feature-based models. In both cases of WV and PS raw imagery-based modeling, ResNet 3D yielded slightly better performance than ResNet 2D. Within hand-crafted feature-based modeling, most of the conventional machine learning algorithms and DNN showed lesser, yet comparable performance. However, when raw bands were combined with spectral vegetation indices (VI) and texture features, there was an increase in the estimation accuracy of all the methods and this was especially true for WV-based modeling. The highest result was found from RFR with WV raw bands, VIs, and texture features (R^2 of 0.87 and RMSE% of 9.71) followed by DNN (R^2 of 0.86 and RMSE% of 10.32) with similar input bands. Overall, the WV-based modeling yielded better predictive performance compared to PS-based models. The ResNet models for soybean did not show overfitting or underfitting for training (Fig. 7a-d).

4.3. Yield prediction for corn

Table 5 compares the performance of both WV and PS based models for corn with respect to R^2 , and RMSE%. Overall, the prediction accuracy for corn was much lower than soybean (Table 5). The highest performance was achieved by the RFR with WV-based raw, VI, and texture bands (R^2 of 0.58 and RMSE% of 9.94). The raw imagery-based ResNet 2D and 3D resulted in a slightly less accurate model. However, the highest model performance within raw imagery-based models was attained by ResNet 3D with PS imageries (R^2 of 0.58 and RMSE% of 9.98). On the other hand, in terms of WV-based raw imageries, ResNet 2D yielded the highest model performance (R^2 of 0.57 and RMSE% of 10.08). According to Fig. 7 (e-h), the ResNet models for corn did not show overfitting or underfitting while training.

4.4. Mapping soybean and corn yield

Predicted grain yield at plot level for soybean and corn fields when using WV and PS data was visualized in Fig. 8. It is worth noting that only the results from the ResNet 3D method which outperformed other methods in most cases were displayed. As shown in Fig. 8a, the distribution pattern of high and low yield values predicted using WV dataset is highly correlated with soybean genotypes, in the case of PS dataset, similar pattern was presented (Fig. 8b).

The corn yield values predicted by ResNet3D models processed with WV and PS datasets demonstrated different spatial distribution patterns (Fig. 8). Further analysis as shown in Fig. 9 revealed that predicted corn yield using WV data were more closely aligned with the actual measured yield, exhibiting a better performance, while PS-based yield values were less consistent with the measured yield curve, particularly for higher or lower values, showing a poorer performance. It is worth noting that, corn fields were exposed to various levels of water stress during the growing season. Water stress at the tasseling and seed filling stage of corn causes permanent yield loss even though water availability is restored in later phenological stages. Thus, to some extent, the various performance of two datasets in corn yield prediction (Fig. 9) is likely attributed to early water stress which complicates corn yield prediction.

5. Discussion

5.1. Feature and imagery based yield prediction comparison and analysis

Crop grain yield prediction using remote sensing imagery-derived features such as reflectance data, VIs, and textures is widely-employed in many previous studies (Maimaitijiang et al. 2020b; Shiu and

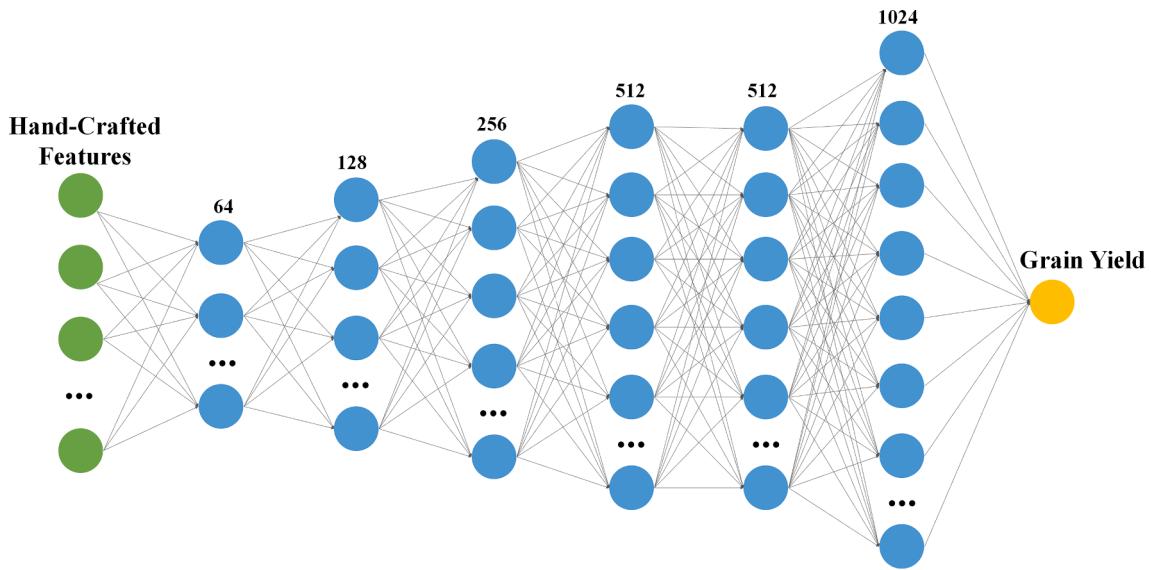


Fig. 3. A schematic illustration of the DNN architecture used for yield prediction in this work.

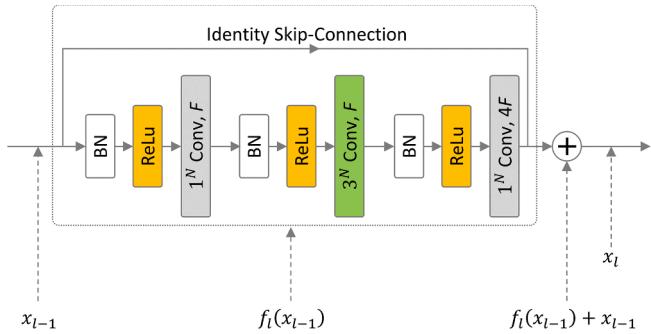


Fig. 4. Architecture of a residual block for 2D ($N = 2$) and 3D ($N = 3$). The convolution (Conv) layers are preceded by one batch normalization (BN) and one rectified linear unit (ReLU) activation function. The kernel size of each Conv layer in 2D and 3D is x^N with $N = 2$ and 3, respectively, where x can be 1 or 3. F is the number of feature maps of the convolutional filter.

Table 3
Input image dimension of WV and PS in ResNet-18 2D and 3D architectures.

Sensors	ResNet 2D	ResNet 3D
WorldView-3 (WV)	$224 \times 224 \times 64$	$224 \times 224 \times 16 \times 4$
PlanetScope (PS)	$224 \times 224 \times 100$	$224 \times 224 \times 4 \times 25$

(Chuang 2019a; Yuan et al. 2019). However, feature-based yield prediction often requires tedious feature extraction from the input dataset, as well as careful feature selection procedures, specifically when the feature space is very large. For example, in our study, the feature space was quite large for both WV and PS imageries, i.e., WV had a larger number of spectral bands with fewer temporal resolution, whereas the PS imageries had larger temporal dimension with fewer spectral bands. In addition to original spectral bands, different VIs and texture features were calculated for each date, which resulted in a large feature space. Additionally, what image features should be extracted for yield prediction from hundreds or even thousands of previously developed VIs, textures and other types of features is also a challenging decision, which requires a high level of domain expertise and training, and tend to be subjective, which might bring potential bias to the prediction results. Moreover, careful feature selection process is often needed after a variety of features were extracted from remote sensing images. In this

study, we employed a random forest-based feature selection method to extract key features from the larger feature space. However, this process is subjective to the specific problem and depends on parameter selection while training the random forest model. Therefore, the feature selection process adds additional uncertainty to the complete handcrafted feature-based yield prediction modeling.

Alternatively, instead of extracting numerous features and averaged to each plot, directly feeding the raw imagery to the deep neural networks for yield prediction allows a rich variety of nonlinear, hierarchical and complex features to be learned from the data itself (not using from previously developed features), which is potentially able to avoid the tedious feature extraction procedure, as well as the subjectivity and uncertainty in feature selection phase (Ball et al. 2017). However, very few studies have been found to utilize direct raw imagery to predict yield from satellite remote sensing, particularly at a fine plot scale within a field (Russello 2018; You et al. 2017). The evolution of deep neural networks has allowed to push the limits of artificial intelligence in many fields, and remote sensing of crops is not an exception. Specifically, different CNN and its variants, such as VGG net (Simonyan and Zisserman 2014), Alex net (Krizhevsky et al. 2012), U-Net (Ronneberger et al. 2015), have proved their efficiency in different remote sensing-based classification, object detection, and segmentation problems. However, the use of CNN architectures in the regression problem has not been explored well for yield prediction. Our study showed the potential of ResNet 2D and 3D architectures in predicting grain yield directly from raw imagery. The advantage of using direct imagery over hand-crafted features is no calculation of VIs or texture information and no feature selection process is required. The ResNet architecture itself performs the feature selection and predicts yield based on the spectral, textural, and temporal information (for ResNet 3D) associated in the imageries.

The usefulness of temporal information was also noticeable in the model results (Table 5, 6). ResNet 3D produced slightly better results in comparison with ResNet 2D for both soybean and corn yield prediction. For example, RMSE% of WV based soybean yield prediction decreased from 9.46% in 2D to 9.18% in 3D, and similarly, for PS imageries, RMSE % decreased from 13.30% in 2D to 12.24% in 3D. Concerning corn yield prediction, for PS based analysis, the RMSE% decreased from 10.20% in 2D to 9.98% in 3D; however, for WV based analysis, the RMSE% did not decrease when ResNet 3D was utilized. Since PS had more data in temporal dimension (25 acquisitions) than WV (4 acquisitions), for both soybean and corn, the yield prediction performance was improved with ResNet 3D. Alternatively, for corn yield prediction with WV imageries,

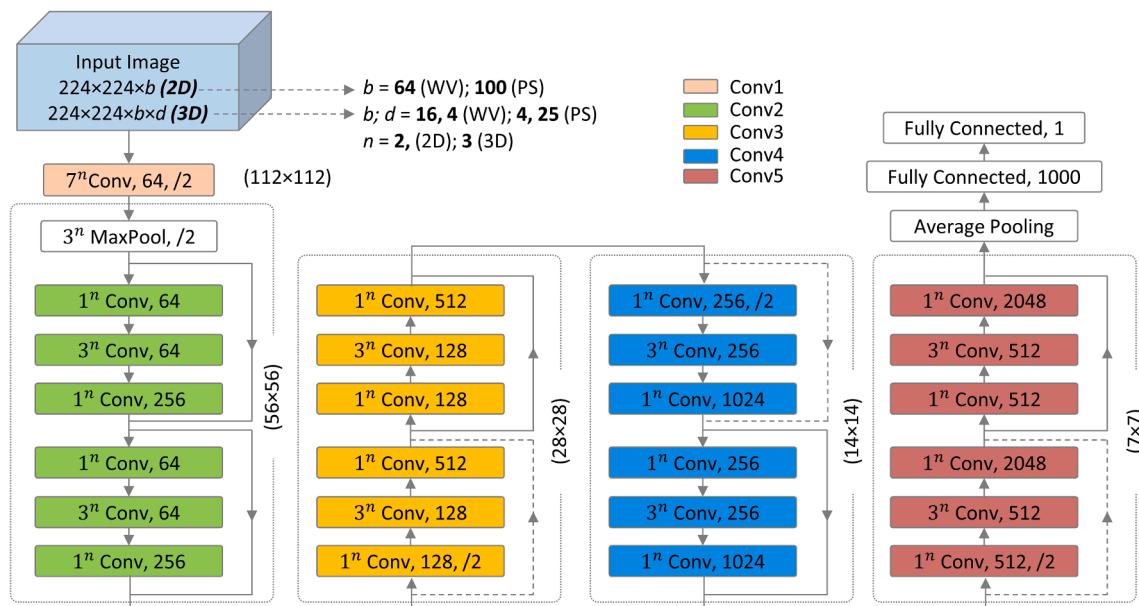


Fig. 5. A schematic illustration of 2D and 3D ResNet-18 architectures used for yield prediction in this study. Each convolutional layer is followed by batch normalization and a ReLU. b is the number of bands in the input image, where $b = 64, 100$ for WorldView-3 (WV), and PlanetScope (PS), respectively for 2D ResNet. However, in 3D ResNet, the bands are divided into b , and d , where d is the temporal dimension. In 3D ResNet, $b = 16, 4$ for WV and PS, respectively, and $d = 4, 25$ for WV and PS, respectively. In Conv layers, n is the dimension of kernel size, where $n = 2, 3$ for 2D, and 3D ResNets, respectively. conv1 spatially down-samples inputs with a spatial stride of 2, and a 3ⁿ max-pooling (MaxPool) with stride 2 is also located before conv2. The output image size is mentioned in parenthesis with each Conv layer.

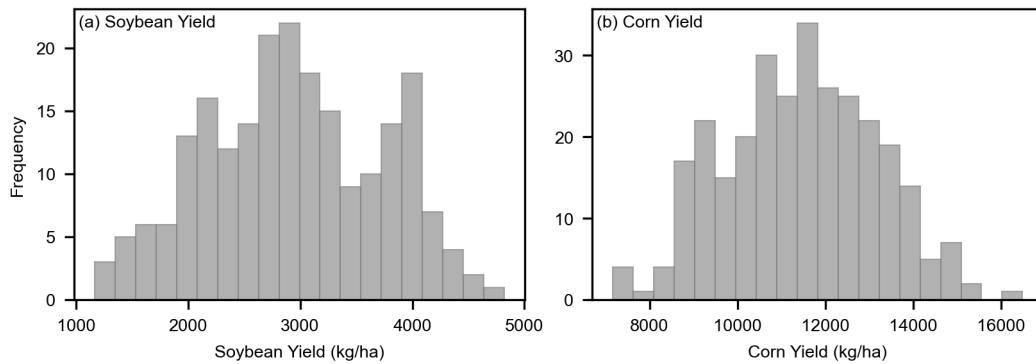


Fig. 6. Histogram of soybean (a) and corn yield (b). The distribution of both crops look normal, however, the overall yield from cornfield is much higher than soybean.

Table 4

Results of soybean yield prediction. Overall, all models provided better results with WV compared to PS. The highest accuracy was achieved from raw imagery-based modeling with ResNet 2D and 3D with WV imageries. A similar scenario was observed from PS imagery as well. ResNet 3D provided a slightly better result than ResNet 2D. By utilizing raw imagery directly in deep residual learning models (ResNet 2D and 3D), it is possible to automatically predict grain yield without any intermediary feature selection.

Satellite	Metrics	Hand-Crafted Features-Based				Imagery-Based	
		PLSR	SVR	RFR	DNN	ResNet 2D	ResNet 3D
WorldView-3 (raw)	R^2	0.63	0.65	0.69	0.74	0.88	0.89
	RMSE%	16.44	16.09	15.06	13.87	9.46	9.18
WorldView-3 (raw + VIs + Texture)	R^2	0.74	0.82	0.87	0.86	/	/
	RMSE%	13.74	11.47	9.71	10.32		
PlanetScope (raw)	R^2	0.55	0.74	0.74	0.79	0.76	0.80
	RMSE%	18.20	13.79	13.83	12.47	13.3	12.24
PlanetScope (raw + VIs + Texture)	R^2	0.62	0.78	0.79	0.84	/	/
	RMSE%	16.67	12.75	12.40	10.96		

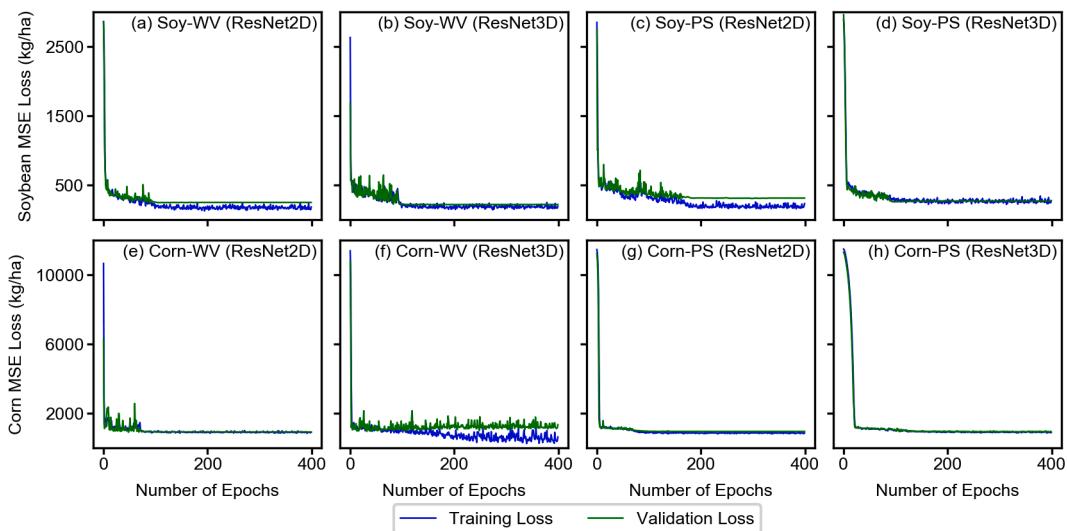


Fig. 7. The training and validation loss curve obtained from each ResNet training process. The model was only saved when the validation loss improved from the previous epoch results. Soybean models (a-d) and Corn models (e-h) did not show overfitting or underfitting throughout the training process.

Table 5

Corn yield prediction results based on WorldView-3 (WV), and PlanetScope (PS) data with different models. The conventional machine learning algorithms and DNN are shown from raw bands, VI, and texture-based modeling, whereas ResNet 2D and 3D are from raw imagery-based modeling. Overall, WV-based deep convolutional neural network modeling performed best utilizing raw imagery as the direct input.

Satellite	Metrics	Hand-Crafted Features-Based				Imagery-Based	
		PLSR	SVR	RFR	DNN	ResNet 2D	ResNet 3D
WorldView-3 (raw)	R^2	0.49	0.51	0.47	0.55	0.57	0.57
WorldView-3 (raw + VIs + Texture)	R^2	0.48	0.56	0.58	0.55	/	/
PlanetScope (raw)	R^2	0.40	0.37	0.42	0.45	0.56	0.58
PlanetScope (raw + VIs + Texture)	R^2	0.47	0.50	0.50	0.49	10.20	9.98
	RMSE%	11.87	12.25	11.69	11.45		
	R^2	11.16	10.85	10.91	11.04		

the performance did not increase with ResNet 3D because the WV had only 4 temporal acquisitions. Additionally, we tested the usefulness of images collected at different growth stages for soybean by training different combinations of images from different dates using ResNet 2D architecture (Fig. 10). Performance metrics from Fig. 10 suggests that if the ResNet 2D model was trained with images collected at later growth stages, the model tends to estimate the crop yield better. For instance, the model trained with only the image collected at the last growth stage (September 4th, 2017) yielded the best performance (R^2 of 0.85 and RMSE% of 10.48) in terms of individual temporal datasets for Soybean. That's why the evaluation result was the highest (R^2 of 0.88 and RMSE% of 9.45) when all four images were combined for Soybean.

5.2. Model performances with WV and PS data for yield prediction

Despite the limited number of image availability, ResNet-3D processed with WV images with more bands and higher spatial resolution yielded better estimation accuracy than PS images, which had fewer bands with lower spatial resolution. The superior performance of the deep learning model with WV can be attributed to the additional bands and better spatial resolution, which provide more accurate information about plant traits (Fig. 11). Many previous studies have emphasized the significance of spectral and spatial information in crop monitoring (Geipel et al. 2014; Mariotto et al. 2013; Sagan et al. 2019). In the VNIR region of WV, yellow and red edge bands, which are missing from PS images, are the two important bands proven to be effective in numerous studies. Yellow band has shown strong sensitivity to subtle changes in

chlorophyll, nitrogen and water status, thus plant health (Maimaitiyiming et al. 2017; Zengeya et al. 2013). Red edge band, on the other hand, has contributed to the improvement of photosynthetic pigments, LAI, biomass and yield estimation at high level of vegetation coverage by solving saturation problem to some degree (Mutanga et al. 2012; Prey and Schmidhalter 2019; Sibanda et al. 2017; Zhou et al. 2017; Zhu et al. 2017). High prediction accuracy of the models in this study may also have benefited from eight additional bands of WV imagery in the SWIR region, which is associated with plant water content, cellulose, and lignin (Kokaly et al. 2009; Kokaly and Clark 1999). More specific significances of red edge and SWIR bands for model prediction will be discussed in the following sections.

Insufficient spatial resolution increases the influence of soil or crop residue, which in turn negatively affects the performance of individual bands and calculated vegetation indices (Du and Noguchi 2017). ResNet 2D, designed to take advantage both spectral and spatial information contained within WV imagery, slightly reduced the prediction accuracy, albeit negligible, and this might be due to the mixed pixel issue (single-pixel containing ensemble spectrum of exposed soil, weed or neighboring plot) caused by 7.2 m resolution SWIR bands.

Results showed that the ResNet-3D model processed with 25 PS images that contain fewer spectral bands and lower spatial resolution produced comparable prediction accuracy to that of WV images. Additionally, taking advantage of high temporal resolution (less than a day revisit time) of PS images, ResNet 3D improved the yield prediction accuracy compared to ResNet 2D. Consistent with previous studies, this finding demonstrates that satellite observations with high temporal

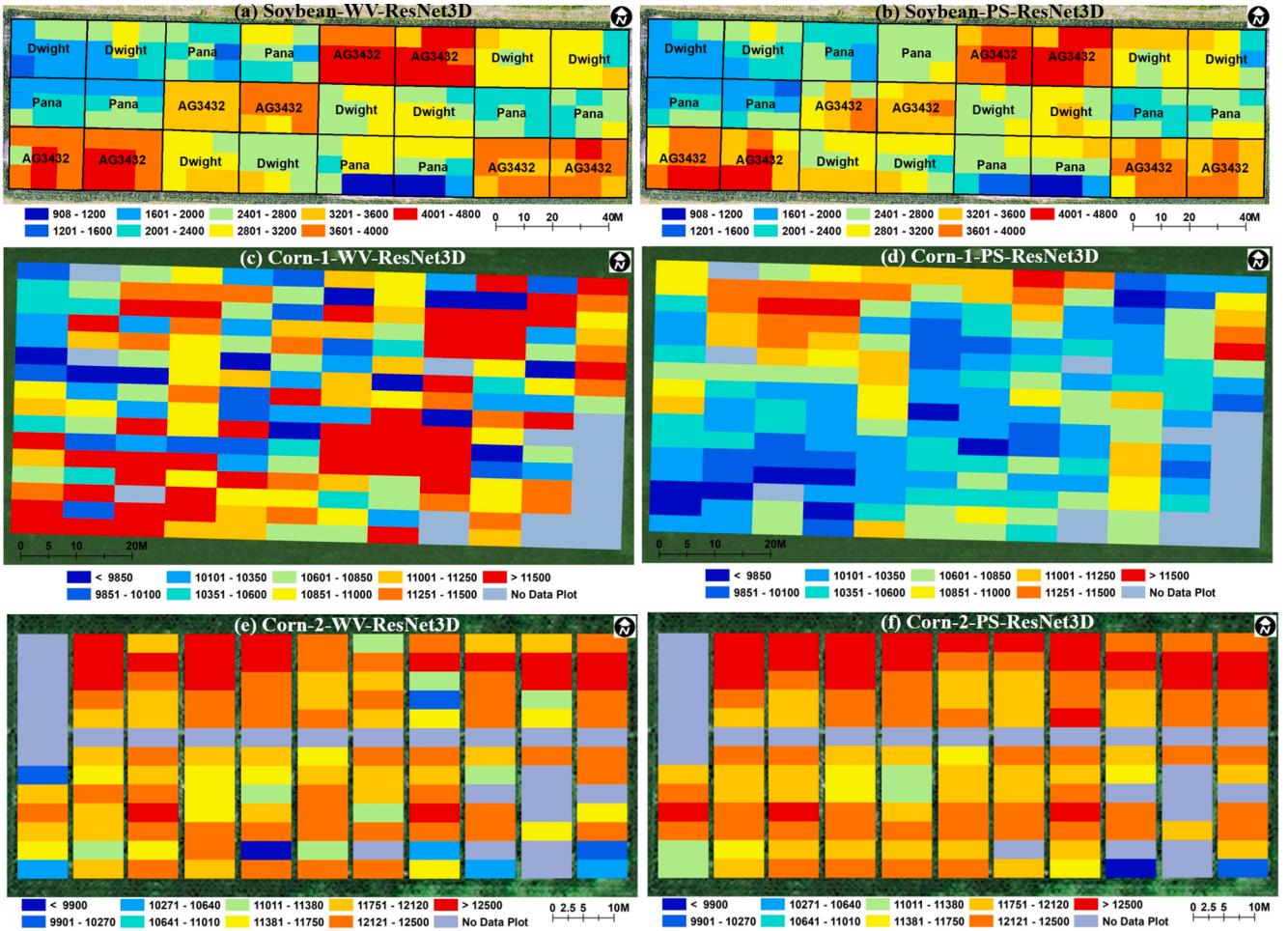


Fig. 8. Spatial distribution of predicted soybean grain yield using ResNet based on WorldView-3 and PlanetScope data. The unit of the values is in kg ha^{-1} .

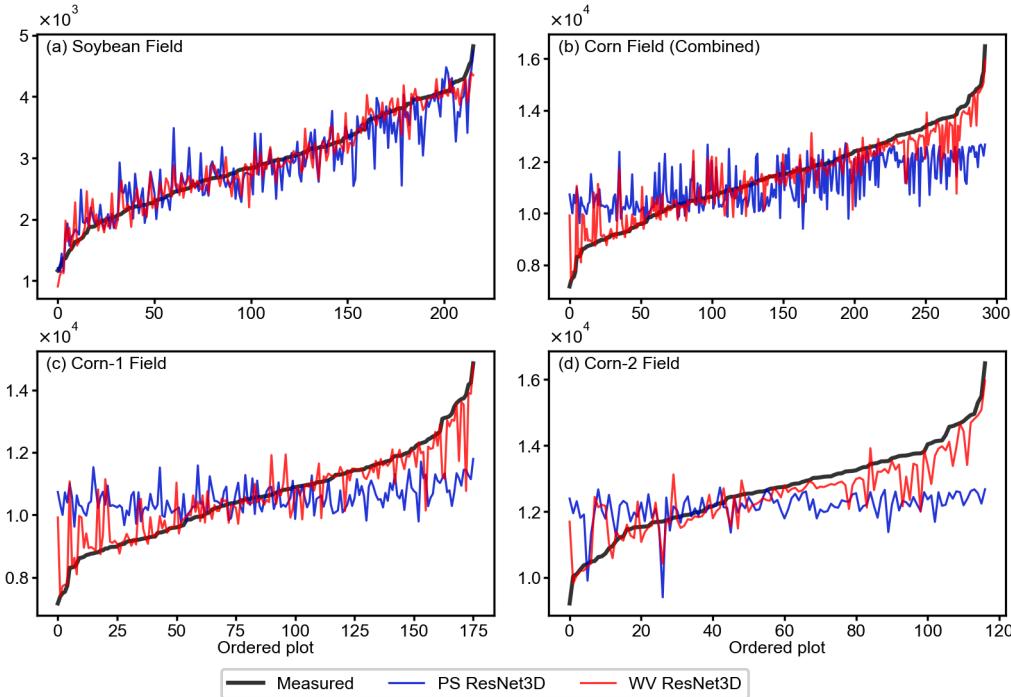


Fig. 9. Comparison of WorldView-3 (WV) and PlanetScope (PS) based yield prediction results at plot level for (a) soybean field, (b) cornfields combined, (c) corn-1 field, and (d) corn-2 field. The predicted yield values from WV (red lines) is more aligned with measured yield, exhibiting a better performance, while PS-based yield values (blue line) are less consistent with the measured yield curve, particularly for higher or lower values, showing a poorer performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resolution provide more useful information about crop growth by capturing critical phenological changes (Ma et al. 2001; Wang et al. 2014; Zhou et al. 2017).

Although the ResNet-3D deep learning model applied to WV images showed promising results, it is worth noting that WV data is not freely available, and frequent cloud cover over cropping regions further reduces the number of cloud-free images available for users due to longer revisit time of WV satellites. On the other hand, PS data are freely available with multiple global coverage per day. It should be noted that a deep learning model that is specifically suited to take advantage of temporal dimension of PS data (e.g., recurrent neural networks) or a self-attention based model may lead to a better performance with PS data compared to WV data.

5.3. Contribution of red-edge band for yield prediction

As demonstrated by Table 4 and 5, a limited number of WV data (only four) collected from the beginning pod stage (R3 stage) to the seed-filling stage (R5) provided significantly higher accuracy for yield prediction compared to 25 PS images representing entire growth stages. This may be attributed to several factors including the higher spatial resolution of WV data and availability of red-edge and SWIR bands that are unique to WV data. Therefore, the significance of red-edge and SWIR bands were evaluated by eliminating one or the other for each model run. It is worth noting that we performed this analysis only for the soybean field due to its higher canopy closure during the image acquisitions which is helpful isolating the effects of red-edge and SWIR bands without the impact of background soil.

The red-edge is defined as the maximum value of the first derivative spectra ($dR/d\lambda$) in the red-edge region (680 nm to 750 nm) (Horler et al. 1983), which has been found to be useful for early stress detection (Horler et al. 1983), estimations of foliar N (Clevers and Gitelson 2013; Li et al. 2014a), chlorophyll concentration (Clevers and Gitelson 2013) and corn yield prediction (Torino et al. 2014). The absence of a red-edge band seemed to reduce the prediction accuracy of ResNet 2D (Fig. 12 and Table 6), i.e., R^2 values dropped from 0.88 to 0.87 and RMSE% increased from 9.46% to 9.69%, respectively. However, a slight improvement was found for ResNet 3D with the exclusion of the red-edge band from the model, possibly because there were only four red-edge bands that were very small compared to the entire data cube consisting of total 64 bands. Our results indicated that the red-edge band has a negligible effect on predicting soybean yield at the early growth stage. Several recent studies found red-edge indices to be more effective in predicting rice grain yield (Kanke et al. 2016; Zhou et al. 2017). However, the results of these studies are not comparable to our findings as they used linear statistical regression between handheld spectroradiometer or UAV-based multispectral imagery and grain yield, which does not account for the non-linear relationship.

5.4. Contribution of SWIR bands of WV imagery for yield prediction

Removal of the SWIR bands showed a noticeable drop in R^2 values

and an increase in RMSE% of yield prediction accuracy. This is especially true for ResNet 3D (R^2 decreased from 0.89 to 0.87 and RMSE% increased from 9.18% to 9.63%, respectively) likely due to the reduction of the data dimension by 50%. This finding confirms previous studies that found SWIR-based indices were stronger predictors of grain yield for wheat (El-Hendawy et al. 2019; Tuvendorj et al. 2019) and corn (Meng et al. 2014). Although one previous work (Galvao et al. 2009) noted that SWIR/VIS band ratios (2250/570 nm) provided the largest positive correlations with grain yield when images were taken at forward scattering direction, the relationship did not hold for images from backward scattering view; there have been no systematic studies on the contribution of SWIR for soybean yield prediction.

SWIR spectrum from 1100 to 2500 nm wavelengths are strongly absorbed by leaf or canopy water content (Ceccato et al. 2001; Ghulam et al. 2007; Ghulam et al. 2008; Kokaly et al. 2009; Kumar et al. 2002) and therefore enables better characterization of crop growth under various environmental stress conditions. This is particularly noteworthy as crops are sensitive to water stress (Nesmith and Ritchie 1992), especially at the emergence, pollination, or grain filling, which may affect the terminal yield (Mkhabela et al. 2010; Yang et al. 2018). However, the improvement of yield prediction accuracy by adding SWIR was not as significant as we expected, which may be attributed to (1) the fact that SWIR bands have lower ground sampling distance (7.2 m) compared to panchromatic band (0.31 m) and VNIR (1.24 m) therefore were not able to capture within plot canopy heterogeneity, and (2) the absence of irrigation effects among the plots of the soybean field. The field received timely rainfall throughout most of the growing season and the soybean varieties were exposed to no water stress (Maimaitijiang et al. 2020b).

These observations support the following conclusions that VNIR data can provide yield estimations for soybean comparable to results of VNIR and SWIR data combined therefore including SWIR for soybean yield prediction is helpful but can be insignificant when there is no water stress. However, crops are sensitive to water availability and yield prediction is crop type dependent. Including SWIR may be helpful for yield prediction at a heterogeneous landscape with different crop types with varying exposure to water deficiency as studies demonstrated that SWIR bands are beneficial to discriminate weeds, soil, crop-soil mix, and crops in late developmental stages (Sidike et al. 2019).

5.5. Performance of different methods across varieties

The adaptability of yield prediction models across different crop varieties is of great importance in terms of field-based high throughput plant phenotyping (Li et al. 2014b; Maimaitijiang et al. 2020b). Thus, the predicting capability and reliability of different models across three soybean varieties were further investigated.

Fig. 13 presented the prediction results based on the WV and PS datasets using hand-crafted features and imagery-based methods, respectively. Overall, except for PLSR, both feature and imagery-based methods presented stable adaptability with decent accuracies across the three varieties with the WV dataset. In the case of PS dataset, feature-based DNN method yielded the most stable adaptability across the three soybean varieties, yet imagery-based methods ResNet 2D and ResNet 3D presented decent results to some extent. It is worth noting that higher accuracies were achieved for genotypes 'Dwight' and 'Pana' compared to 'AG3432' regardless of modeling methods and datasets, 'AG3432' is more different from other two genotypes in terms of canopy structure characteristics with denser and higher canopy, as well as higher grain yield production, which would also cause optical saturation issue, and potentially leads to lower yield prediction accuracy (Maimaitijiang et al. 2020b; Vergara-Díaz et al. 2016). With respect to different methods, PLSR exhibited the poorest performance. PLSR could often deal with collinearity issues, however, due to the limitation of characterizing the complex and non-linear relationship between hand-crafted canopy spectral/texture features and crop grain yield, and likely led to poorer

Table 6
Validation statistics of Resnet-18 models for soybean yield prediction using WorldView-3 imagery.

Satellite	NO. of bands	Metrics	Imagery based	
			ResNet 2D	ResNet 3D
WorldView-3 (including all bands)	64	R^2	0.88	0.89
		RMSE%	9.46	9.18
WorldView-3 (excluding red edge band)	60	R^2	0.87	0.89
		RMSE%	9.69	9.12
WorldView-3 (excluding SWIR bands)	32	R^2	0.87	0.87
		RMSE%	9.72	9.63

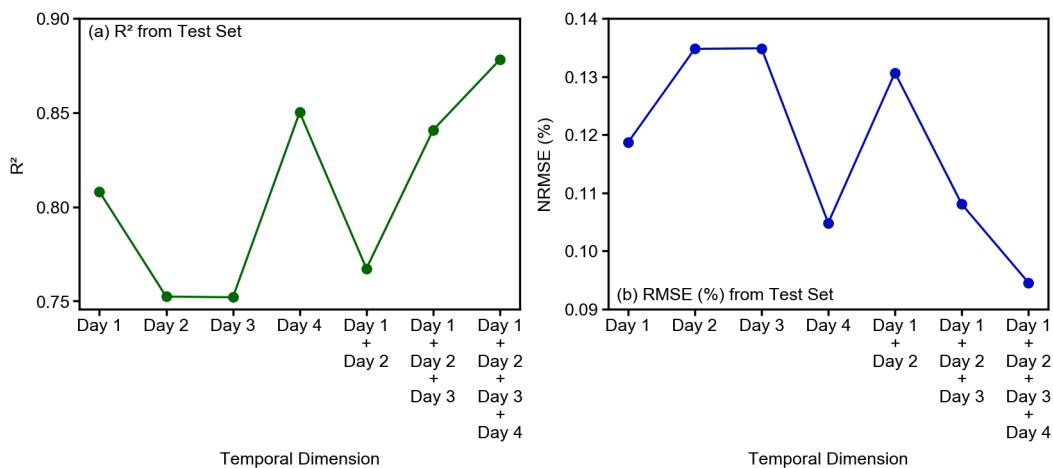


Fig. 10. Model evaluation metrics for testing the temporal effect of WorldView-3 (WV) images with ResNet 2D models for Soybean. R^2 (a) and RMSE% (b) were calculated from the testing set of Soybean. The X-axis represents different input configurations for model training with images of different dates. Day 1, 2, 3, and 4 represent images collected on July 16th, July 22nd, August 9th, and September 4th of 2017, respectively.

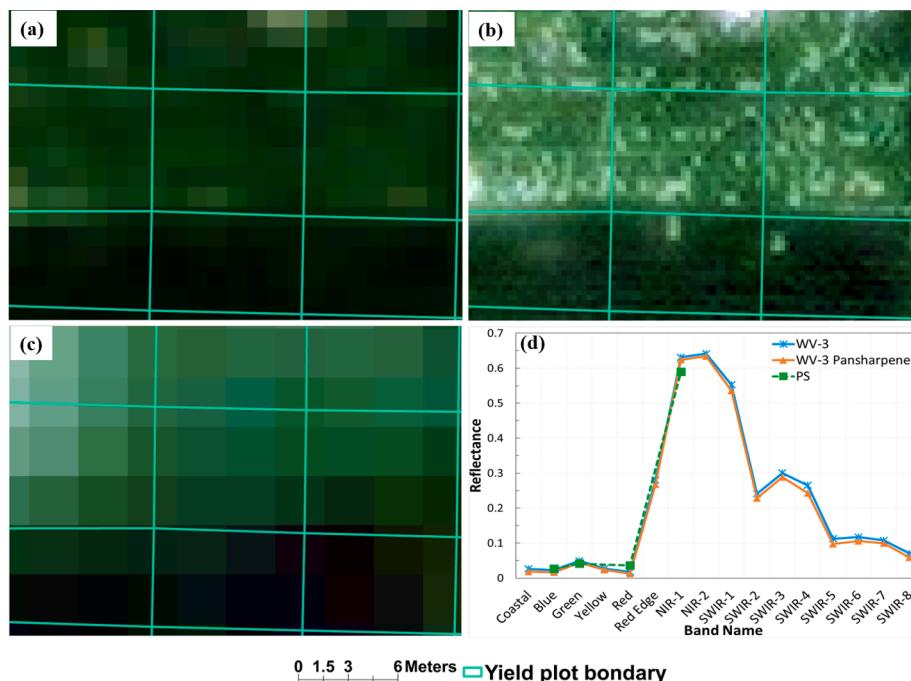


Fig. 11. Comparison of Worlview-3 (WV) and PlanetScope (PS) data. Yield plot overlaid with WV image (a), Pan sharpened WV image (b), and PS image (c). Spectral profiles of soybean from Wroldveiw-3, Pan sharpened WV and PS data (d). WorldView-3 data has higher spatial resolution and suffers less mixed pixels among neighboring yield plots. In addition, Pan sharpening of WV data has an ignorable effect on the spectral properties. Moreover, WV data covers larger spectral range (i.e., red edge and SWIR, etc.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

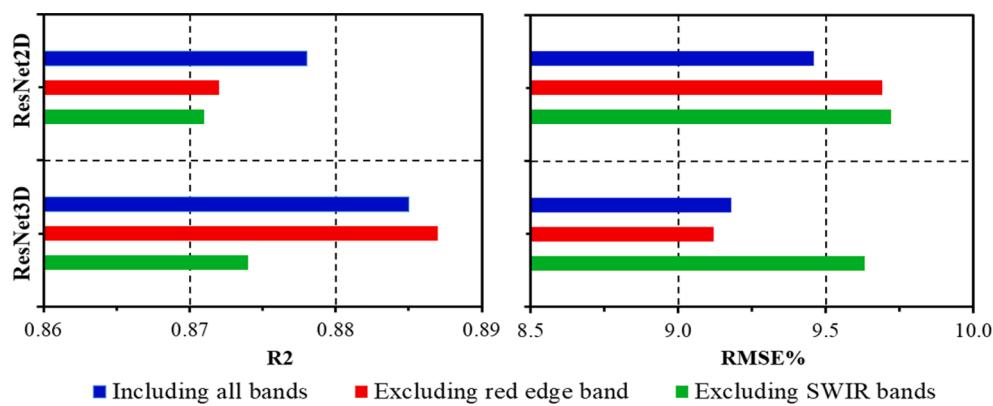


Fig. 12. Comparison of prediction results in the case of using all bands, excluding the red-edge band, and excluding the SWIR bands. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

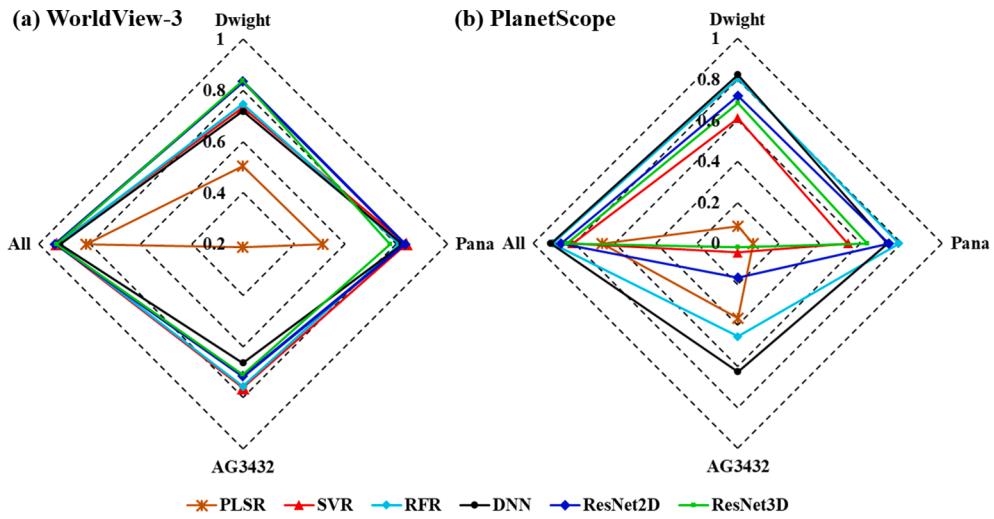


Fig. 13. Comparison of R^2 for soybean grain yield prediction using different modeling methods. R^2 in the case using the WV dataset (a), R^2 in the case using the PS dataset (b).

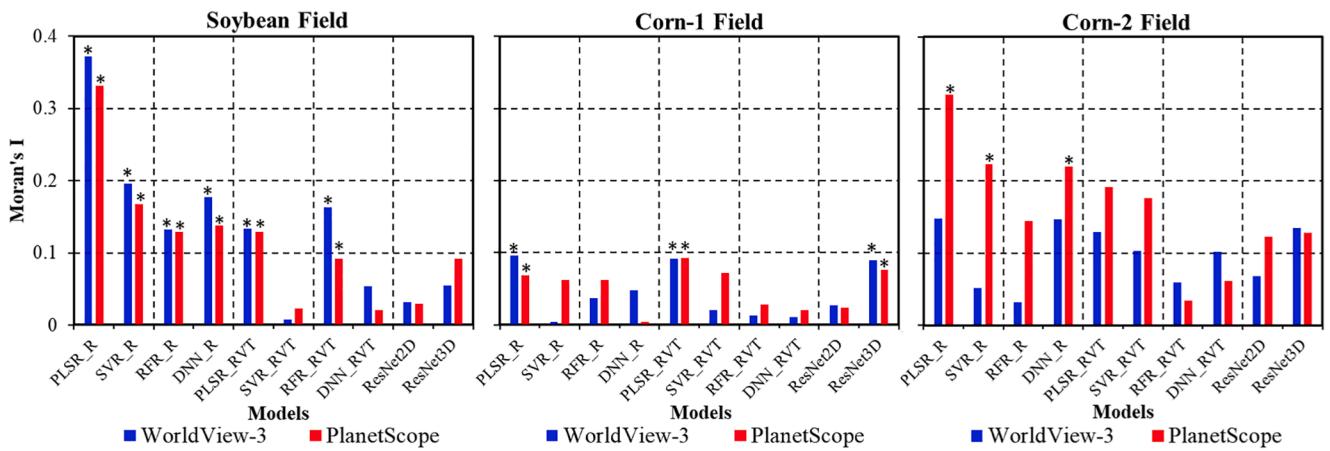


Fig. 14. Comparison of Moran's I values of different models for soybean, corn-1 and corn-2 fields. The “*” on top of each bar means the Moran's I is statistically significant at 0.01P-value level. (*PLSR_R, SVR_R, RFR_R and DNN_R are models using only raw bands; PLSR_RVT, SVR_RVT, RFR_RVT and DNN_RVT are models using raw bands, vegetation indices and texture features.)

adaptability for yield prediction across different soybean varieties (Maimaitijiang et al. 2020a; Wang et al. 2018). Nonetheless, the applicability of imagery-based yield prediction, particularly concerning plant phenotyping, should be evaluated across a larger number of varieties, as well as different crop types.

5.6. Performance of different methods over space

To investigate the adaptability of grain yield prediction models over space, Global Moran's I (GMI), a measure of spatial autocorrelation demonstrated by randomness of error (Anselin 1995), was computed for plot-level grain yield prediction errors from each model in soybean, corn-1, and corn-2 fields separately in this study (Fig. 14).

As shown in Fig. 14, in the case of soybean yield prediction, statistically significant weak to moderate spatial clustering patterns were noticed when using only raw bands as input variables (PLSR_R, SVR_R, RFR_R and DNN_R) for both WV and PS datasets. Inclusion of vegetation indices and texture features to raw bands of WV and PS data (denoted as PLSR_RVT, SVR_RVT, RFR_RVT and DNN_RVT) led to the decrease GMI values (0 is the best), indicating improved performance feature-based methods over space. For both soybean and corn yield prediction, ResNet 2D and ResNet 3D showed random distribution patterns,

indicating stronger generalizability over space. It is worth noting that the PLSR method demonstrated a significantly stronger spatial clustering pattern compared to other methods, which is likely due to its limitation in dealing with complex and non-linear relationships, and resulted in poorer adaptability over space (Maimaitijiang et al. 2020b). Future research should be conducted over a larger region with higher spatial complexity and variations.

6. Conclusion

Our work presented one of the very first studies on using satellite imagery for field-scale yield prediction using 2D and 3D deep residual network architectures with raw spectral imagery as the direct input. The ResNet architecture implemented in this paper demonstrated that an imagery-based deep learning approach can fully utilize spectral, spatial, temporal information contained in the satellite data eliminating the need for feature engineering, selection, and manipulation. The results of direct estimation were compared with feature-based machine learning approaches including PLSR, SVR, and deep neural networks. Our results demonstrated the following.

- (1) Direct imagery-based deep ResNet outperformed feature-fed deep learning for yield prediction in many cases investigated in this study. Percent root mean square error was about 10% regardless of the crop types and irrigation conditions, which explained nearly 90% variance in field-scale yield. The performance of direct imagery-based deep learning was comparable, if not superior, to feature-fed deep learning approaches for different crop varieties, indicating the potential of fully automated, AI-driven agriculture providing yield predictions down to crop varieties among the same species.
- (2) Even with the limited number of observations, ResNet3D model processed with WV data may provide reliable predictions of yield over the model derived from multi-temporal PS satellite data representing the entire growing season due to its higher spatial resolution and availability of RedEdge and SWIR bands. Randomness in error measured by Moran's I metrics, demonstrated that the deep learning model with both WV and PS data are capable of providing yield models that are stable, insensitive to crop types and species, and scalable over time and space. However, the model with WV data consistently outperformed the model with PS in spatial adaptability with lower Moran's I in most of the cases, especially in predicting high and low yield scenarios.
- (3) The inclusion of red-edge and SWIR bands improved the prediction accuracy. However, the improvement was not significant as expected since there were only 4 red-edge bands that were limited compared to the other bands of WV data; SWIR suffers from mixed pixel issues due to its lower spatial resolutions.
- (4) High spatial resolution satellite data (WV and PS) can provide reliable estimates of field-scale yield, which could significantly improve food security by real-time and efficient agricultural management.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Science Foundation (IAA-1355406 and IIA-1430427), and in part by the National Aeronautics and Space Administration (NNX15AK03H). The authors thank the aboveground biomass and grain yield harvest team from University of Missouri in Columbia who spent long and strenuous hours collecting data. The authors also would like to thank the editor and the anonymous reviewers for their thoughtful review and constructive comments.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M., 2016. Tensorflow: A system for large-scale machine learning. In, 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283.
- Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H.S., Radom, S., 2018. Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 4563–4577.
- Al-Gaadi, K.A., Hassaballa, A.A., Tola, E., Kayad, A.G., Madugundu, R., Alblewi, B., Assiri, F., 2016. Prediction of Potato Crop Yield Using Precision Agriculture Techniques. *PLoS ONE* 11.
- Anselin, L., 1995. Local Indicators of Spatial Association - Lisa. *Geograp. Anal.* 27, 93–115.
- Ball, J.E., Anderson, D.T., Chan, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *J. Appl. Remote Sens.* 11, 042609.
- Baret, F., Guyot, G., 1991. Potentials and Limits of Vegetation Indexes for Lai and Apar Assessment. *Remote Sens. Environ.* 35, 161–173.
- Batchelor, W.D., Basso, B., Paz, J.O., 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *Eur. J. Agron.* 18, 141–158.
- Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sens. Environ.* 184, 668–681.
- Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 114, 1312–1323.
- Betbeder, J., Fieuzal, R., Baup, F., 2016. Assimilation of LAI and Dry Biomass Data From Optical and SAR Images Into an Agro-Meteorological Model to Estimate Soybean Yield. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9, 2540–2553.
- Biganzoli, E., Boracchi, P., Mariani, L., Marubini, E., 1998. Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach. *Stat. Med.* 17, 1169–1186.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS* 114, 2189–2194.
- Cai, R.H., Yu, D.L., Oppenheimer, M., 2014. Estimating the Spatially Varying Responses of Corn Yields to Weather Variations using Geographically Weighted Panel Regression. *J. Agric. Resour. Econom.* 39, 230–252.
- Cai, Y.P., Guan, K.Y., Lobell, D., Potgieter, A.B., Wang, S.W., Peng, J., Xu, T.F., Asseng, S., Zhang, Y.G., You, L.Z., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* 274, 144–159.
- Cai, Y.P., Guan, K.Y., Peng, J., Wang, S.W., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47.
- Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., Gregoire, J.M., 2001. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sens. Environ.* 77, 22–33.
- Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* 23, 344–351.
- Cortes, C., Vapnik, V., 1995. Support-Vector Networks. *Mach. Learn.* 20, 273–297.
- Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74, 229–239.
- Deering, D., 1975. Measuring“ forage production” of grazing units from Landsat MSS data. In: Proceedings of the Tenth International Symposium of Remote Sensing of the Environment, pp. 1169–1198.
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sens. Environ.* 92, 548–559.
- Du, M.M., Noguchi, N., 2017. Monitoring of Wheat Growth Status and Mapping of Wheat Yield's within-Field Spatial Variations Using Color Images Acquired from UAV-camera System. *Remote Sens.* 9.
- El-Hendawy, S.E., Al-Suhaimani, N.A., Elsayed, S., Hassan, W.M., Dewir, Y.H., Refay, Y., Abdella, K.A., 2019. Potential of the existing and novel spectral reflectance indices for estimating the leaf water status and grain yield of spring wheat exposed to different irrigation rates. *Agric. Water Manag.* 217, 356–373.
- Fieuzal, R., Sicre, C.M., Baup, F., 2017. Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *Int. J. Appl. Earth Obs. Geoinf.* 57, 14–23.
- Galvao, L.S., Roberts, D.A., Formaggio, A.R., Numata, I., Breunig, F.M., 2009. View angle effects on the discrimination of soybean varieties and on the relationships between vegetation indices and yield using off-nadir Hyperion data. *Remote Sens. Environ.* 113, 846–856.
- Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266.
- Geipel, J., Link, J., Claupein, W., 2014. Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System. *Remote Sens.* 6, 10335–10355.
- Georganos, S., Grappa, T., Vanhuyse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E., 2018. Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application (vol 55, pg 221, 2017). *Gisci. Remote Sens.*, 55, Iv–Iv.
- Grevrey, M., Dimopoulos, I., Lek, S., 2006. Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecol. Model.* 195, 43–50.
- Ghulam, A., Li, Z.L., Qin, Q.M., Tong, Q.X., Wang, J.H., Kasimu, A., Zhu, L., 2007. A method for canopy water content estimation for highly vegetated surfaces-shortwave infrared perpendicular water stress index. *Sci. China Ser. D-Earth Sci.* 50, 1359–1368.
- Ghulam, A., Li, Z.L., Qin, Q.M., Yimit, H., Wang, J.H., 2008. Estimating crop water stress with ETM plus NIR and SWIR data. *Agric. For. Meteorol.* 148, 1679–1695.
- Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* 161, 165–173.
- Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160, 271–282.
- Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80, 76–87.
- Gitelson, A.A., Merzlyak, M.N., 1997. Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* 18, 2691–2697.
- Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32.

- Gomez, D., Salvador, P., Sanz, J., Casanova, J.L., 2019. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote Sens.* 11.
- Guide, E.U.S., 2009. Atmospheric correction module: QUAC and FLAASH user's guide. Version 4, 1–44.
- Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81, 416–426.
- Hara, K., Kataoka, H., Satoh, Y., 2018. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imangenet?. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 6546–6555.
- Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. *IEEE Trans. Syst., Man, Cybernet.* 610–621.
- Hariharan, S., Mandal, D., Tirolkar, S., Kumar, V., Bhattacharya, A., Lopez-Sanchez, J. M., 2018. A Novel Phenology Based Feature Subset Selection Technique Using Random Forest for Multitemporal PolSAR Crop Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 4244–4258.
- Harries, K., 2006. Extreme spatial variations in crime density in Baltimore County, MD. *Geoforum* 37, 404–416.
- Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., & Carron, J., 2019. Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors*, 19.
- He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. In: European conference on computer vision. Springer, pp. 630–645.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Horler, D., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4, 273–288.
- Houborg, R., McCabe, M.F., 2018. Daily Retrieval of NDVI and LAI at 3 m Resolution via the Fusion of CubeSat, Landsat, and MODIS Data. *Remote Sens.* 10.
- Huang, J.X., Ma, H.Y., Sedano, F., Lewis, P., Liang, S., Wu, Q.L., Su, W., Zhang, X.D., Zhu, D.H., 2019. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model. *Eur. J. Agron.* 102, 1–13.
- Huang, J.X., Tian, L.Y., Liang, S.L., Ma, H.Y., Becker-Reshef, I., Huang, Y.B., Su, W., Zhang, X.D., Zhu, D.H., Wu, W.B., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Imran, M., Stein, A., Zurita-Milla, R., 2015. Using geographically weighted regression kriging for crop yield mapping in West Africa. *Int. J. Geograp. Inform. Sci.* 29, 234–257.
- Ines, A.V.M., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164.
- Ioffe, S., & Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
- Jiang, H., Hu, H., Zhong, R.H., Xu, J.F., Xu, J.L., Huang, J.F., Wang, S.W., Ying, Y.B., Lin, T., 2019. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Glob. Change Biol.*
- Jiang, Z.Y., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112, 3833–3845.
- Jin, Z.N., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128.
- Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* 141, 116–128.
- Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bedard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* 218, 74–84.
- Kanke, Y., Tubana, B., Dalen, M., Harrell, D., 2016. Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precis. Agric.* 17, 507–530.
- Kern, A., Barcza, Z., Marjanovic, H., Arendas, T., Fodor, N., Bonis, P., Bognar, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agric. For. Meteorol.* 260, 300–320.
- Khaki, S., Wang, L.Z., 2019. Crop Yield Prediction Using Deep Neural Networks. *Front. Plant Sci.* 10.
- Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agric.* 153, 213–225.
- Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., Wessman, C.A., 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sens. Environ.* 113, S78–S91.
- Kokaly, R.F., Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67, 267–287.
- Kowalik, W., Dabrowska-Zielinska, K., Meroni, M., Raczkiewicz, T.U., de Wit, A., 2014. Yield estimation using SPOT-VEGETATION products: A case study of wheat in European countries. *Int. J. Appl. Earth Obs. Geoinf.* 32, 228–239.
- Krizhevsky, A., Sutskever, I., & Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105.
- Kumar, L., Schmidt, K., Dury, S., Skidmore, A., 2002. Imaging spectrometry and vegetation science. In: Imaging spectrometry. Springer, pp. 111–155.
- Kuwata, K., Shibasaki, R., 2016. Estimating corn yield in the United States with modes EVI and machine learning methods. In: ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, p. 3.
- Langkvist, M., Karlsson, L., Loutfi, A., 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recogn. Lett.* 42, 11–24.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Li, F., Miao, Y.X., Feng, G.H., Yuan, F., Yue, S.C., Gao, X.W., Liu, Y.Q., Liu, B., Ustine, S. L., Chen, X.P., 2014a. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Res.* 157, 111–123.
- Li, L., Zhang, Q., Huang, D., 2014b. A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111.
- Li, Y., Zhou, Q.G., Zhou, J., Zhang, G.F., Chen, C., Wang, J., 2014c. Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecol. Model.* 291, 15–27.
- Ma, B.L., Dwyer, L.M., Costa, C., Cober, E.R., Morrison, M.J., 2001. Early prediction of soybean yield from canopy reflectance measurements. *Agron. J.* 93, 1227–1234.
- Ma, J.W., Lee, K., Choi, K.Y., Heo, J., 2017. Rice Yield Estimation of South Korea from Year 2003–2016 Using Stacked Sparse AutoEncoder. *Korean J. Remote Sens.* 33, 631–640.
- Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A.M., Erkholi, H., Fritsch, F.B., 2020a. Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning. *Remote Sens.* 12, 1357.
- Maimaitijiang, M., Sagan, V., Sidike, P., Esposito, F., Fritsch, F.B., 2020b. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* 237, 111599.
- Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K.T., Maw, M.J.W., Shakoor, N., Mockler, T., Fritsch, F.B., 2019. Vegetation Index Weighted Canopy Volume Model (CVWV) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. *ISPRS J. Photogramm. Remote Sens.* 151, 27–41.
- Maimaitiyiming, M., Ghulam, A., Bozzolo, A., Wilkins, J.L., Kwasniewski, M.T., 2017. Early Detection of Plant Physiological Responses to Different Levels of Water Stress Using Reflectance Spectroscopy. *Remote Sens.* 9.
- Maimaitiyiming, M., Sagan, V., Sidike, P., Kwasniewski, M.T., 2019. Dual Activation Function-Based Extreme Learning Machine (ELM) for Estimating Grapevine Berry Yield and Quality. *Remote Sens.* 11.
- Mariotto, I., Thenkabail, P.S., Huete, A., Slonecker, E.T., Platonov, A., 2013. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HypIRI mission. *Remote Sens. Environ.* 139, 291–305.
- Meng, W., Fu-Lu, T., Wen-Jiao, S., 2014. Corn Yield Forecasting in Northeast China Using Remotely Sensed Spectral Indices and Crop Phenology Metrics. *J. Integr. Agric.* 13, 1538–1545.
- Meng, Z., Li, L.L., Tang, X., Feng, Z.X., Jiao, L.C., Liang, M.M., 2019. Multipath Residual Network for Spectral-Spatial Hyperspectral Image Classification. *Remote Sens.* 11.
- Mkhabela, M., Bullock, P., Gervais, M., Finlay, G., Sapirstein, H., 2010. Assessing indicators of agricultural drought impacts on spring wheat yield and quality on the Canadian prairies. *Agric. For. Meteorol.* 150, 399–410.
- Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* 18, 399–406.
- Nesmith, D.S., Ritchie, J.T., 1992. Effects of Soil Water-Deficits during Tassel Emergence on Development and Yield Component of Maize (*Zea-Mays*). *Field Crops Res.* 28, 251–256.
- Nevaluvi, P., Narra, N., Lipping, T., 2019. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* 163.
- Nichol, J.E., Sarker, M.L.R., 2011. Improved Biomass Estimation Using the Texture Parameters of Two High-Resolution Optical Sensors. *IEEE Trans. Geosci. Remote Sens.* 49, 930–948.
- Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* 121, 57–65.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L., 2019. PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, pp. 8024–8035.
- Penuelas, J., Baret, F., Filella, I., 1995. Semiempirical Indexes to Assess Carotenoids Chlorophyll-a Ratio from Leaf Spectral Reflectance. *Photosynthetica* 31, 221–230.
- Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Obs. Geoinf.* 8, 26–33.
- Prey, L., Schmidhalter, U., 2019. Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat. *ISPRS J. Photogramm. Remote Sens.* 149, 176–187.
- Qian, B., De Jong, R., Gameda, S., 2009. Multivariate analysis of water-related agroclimatic factors limiting spring wheat yields on the Canadian prairies. *Eur. J. Agron.* 30, 140–150.

- Quarmby, N.A., Milnes, M., Hindle, T.L., Silleos, N., 1993. The Use of Multitemporal Ndvi Measurements from Avhrr Data for Crop Yield Estimation and Prediction. *Int. J. Remote Sens.* 14, 199–210.
- Ren, J.Q., Chen, Z.X., Zhou, Q.B., Tang, H.J., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* 10, 403–413.
- Rischbeck, P., Elsayed, S., Mistele, B., Barneier, G., Heil, K., Schmidhalter, U., 2016. Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. *Eur. J. Agron.* 78, 44–59.
- Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55, 95–107.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241.
- Rouse, J., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* 351, 309.
- Russello, H., 2018. Convolutional neural networks for crop yield prediction using satellite images. *IBM Center for Advanced Studies*.
- Sagan, V., Maimaitijiang, M., Sidike, P., Maimaitiyiming, M., Erkbol, H., Hartling, S., Peterson, K., Peterson, J., Burken, J., Fritschi, F.U.A.V., 2019. Satellite multiscale data fusion for crop monitoring and early stress detection. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Shiu, Y.-S., Chuang, Y.-C., 2019a. Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models. *Remote Sens.* 11, 111.
- Shiu, Y.S., Chuang, Y.C., 2019b. Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models. *Remote Sens.* 11.
- Sibanda, M., Mutanga, O., Rouget, M., Kumar, L., 2017. Estimating Biomass of Native Grass Grown under Complex Management Treatments Using WorldView-3 Spectral Derivatives. *Remote Sens.* 9.
- Sidike, P., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Shakoor, N., Burken, J., Mockler, T., Fritschi, F.B., 2019. dPEN: deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. *Remote Sens. Environ.* 221, 756–772.
- Simonyan, K., & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn.* 15, 1929–1958.
- Srivastava, R.K., Greff, K., & Schmidhuber, J., 2015. Training very deep networks. In: *Advances in neural information processing systems*, pp. 2377–2385.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.
- Torino, M.S., Ortiz, B.V., Fulton, J.P., Balkcom, K.S., Wood, C.W., 2014. Evaluation of Vegetation Indices for Early Assessment of Corn Status and Yield Potential in the Southeastern United States. *Agron. J.* 106, 1389–1401.
- Tucker, Compton J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Rem. Sens. Environ.* 8 (2), 127–150.
- Tucker, C., Holben, B., Elgin Jr, J., McMurtrey III, J., 1980. Relationship of spectral data to grain yield variation. *Photogramm. Eng. Remote Sens.* 46, 657–666.
- Tuvendorj, B., Wu, B.F., Zeng, H.W., Batdelger, G., Nanzad, L., 2019. Determination of Appropriate Remote Sensing Indices for Spring Wheat Yield Estimation in Mongolia. *Remote Sens.* 11.
- van Hinsbergen, C.P.I., van Lint, J.W.C., van Zuylen, H.J., 2009. Bayesian committee of neural networks to predict travel times with confidence intervals. *Transport. Res. Part C-Emerg. Technolog.* 17, 498–509.
- Vergara-Díaz, O., Zaman-Allah, M.A., Masuka, B., Hornero, A., Zarco-Tejada, P., Prasanna, B.M., Cairns, J.E., Araus, J.L., 2016. A Novel Remote Sensing Approach for Prediction of Maize Yield Under Different Conditions of Nitrogen Fertilization. *Front. Plant Sci.* 7.
- Vergara-Díaz, O., Zaman-Allah, M.A., Masuka, B., Hornero, A., Zarco-Tejada, P., Prasanna, B.M., Cairns, J.E., Araus, J.L., 2016. A novel remote sensing approach for prediction of maize yield under different conditions of nitrogen fertilization. *Front. Plant Sci.* 7, 666.
- Wang, J., Ding, J., Abulimiti, A., Cai, L., 2018. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy. *Ebinur Lake Wetland, Northwest China. PeerJ* 6, e4703.
- Wang, L.G., Tian, Y.C., Yao, X., Zhu, Y., Cao, W.X., 2014. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. *Field Crops Research* 164, 178–188.
- Wang, P.J., Sun, R., Zhang, J.H., Zhou, Y.Y., Xie, D.H., Zhu, Q.J., 2011. Yield estimation of winter wheat in the North China Plain using the remote-sensing-photosynthesis-yield estimation for crops (RS-P-YEC) model. *Int. J. Remote Sens.* 32, 6335–6348.
- Wang, Y., Zhang, Z., Feng, L., Du, Q., Runge, T., 2020. Combining Multi-Source Data and Machine Learning Approaches to Predict Winter Wheat Yield in the Conterminous United States. *Remote Sens.* 12, 1232.
- Yang, Q., Shi, L.S., Han, J.Y., Zha, Y.Y., Zhu, P.H., 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.* 235, 142–153.
- Yang, Y., Anderson, M.C., Gao, F., Wardlow, B., Hain, C.R., Otkin, J.A., Alfieri, J., Yang, Y., Sun, L., Dulaney, W., 2018. Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE, USA. *Remote Sens. Environ.* 210, 387–402.
- You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. *Thirty-First AAAI Conference on Artificial Intelligence*.
- Yuan, W., Wijewardane, N.K., Jenkins, S., Bai, G., Ge, Y., Graef, G.L., 2019. early prediction of Soybean traits through color and texture features of canopy RGB imagery. *Sci. Rep.* 9, 1–17.
- Zengeya, F.M., Mutanga, O., Murwira, A., 2013. Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape. *Int. J. Appl. Earth Obs. Geoinf.* 21, 513–524.
- Zhou, X., Zheng, H.B., Xu, X.Q., He, J.Y., Ge, X.K., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C., 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* 130, 246–255.
- Zhu, Y.H., Liu, K., Liu, L., Myint, S.W., Wang, S.G., Liu, H.X., He, Z., 2017. Exploring the Potential of WorldView-2 Red-Edge Band-Based Vegetation Indices for Estimation of Mangrove Leaf Area Index with Machine Learning Algorithms. *Remote Sens.* 9.