Detection and Prediction of Nutrient Deficiency Stress using Longitudinal Aerial Imagery

Saba Dadsetan[1,2*], **Gisele Rose**[1], **Naira Hovakimyan**[1,3], **Jennifer Hobbs**[1]
[1]Intelinair, Inc., Champaign, IL 61820
[2]University of Pittsburgh, School of Computing and Information, Pittsburgh, PA 15213
[3]University of Illinois at Urbana Champaign, Department of Mechanical Science and Engineering, Urbana, IL 61801
sabadad@cs.pitt.edu, gisele@intelinair.com, naira@intelinair.com, jennifer@intelinair.com

Abstract

Early, precise detection of nutrient deficiency stress (NDS) has key economic as well as environmental impact; precision application of chemicals in place of blanket application reduces operational costs for the growers while reducing the amount of chemicals which may enter the environment unnecessarily. Furthermore, earlier treatment reduces the amount of yield loss and therefore boosts crop production during a given season. With this in mind, we collect sequences of high-resolution aerial imagery and construct semantic segmentation models to detect and predict NDS across the field; our work sits at the intersection of agriculture, remote sensing, and deep learning. First, we establish a baseline for full-field detection of NDS and quantify the impact of pretraining, backbone architecture, input representation, and sampling strategy. We then quantify the amount of information available at different points in the season by building a single-timestamp model based on a U-Net. Next, we construct our proposed spatiotemporal architecture, which combines a U-Net with a convolutional LSTM to accurately detect regions of the field showing NDS; this approach has an impressive IOU score of 0.53. Finally, we show that this architecture can be trained to *predict* regions of the field which are expected to show NDS in a later flight- potentially more than three weeks in the future- maintaining an IOU score of 0.47-0.51 depending on how far in advance the prediction is made. We will also release a dataset which we believe will benefit the computer vision, remote sensing, and agriculture fields. This work contributes to the recent developments in deep learning for remote sensing and agriculture while addressing a key social challenge with implications for economics and sustainability.

Introduction

Precision agriculture is a key rising area of interest for the application of deep learning approaches. Computer vision approaches and applications in agriculture simultaneously address key social needs while furthering our understanding of the machine learning field by addressing unique theoretical and computational challenges.

Precision agriculture and sustainable practices are central to addressing challenges around economic hardship, food and clean water scarcity, and climate change (Rolnick et al. 2019). The Food and Agriculture Organization (FAO) of the United Nations estimates that to feed the world's ever-growing population 50% more food needs to be produced by 2050 (FAO et al. 2017). This demand for an increase in production comes amidst challenges brought about by environmental changes as agriculture is both impacted by and contributes to climate change and other environmental issues.

Central to these challenges is the task of identifying Nutrient Deficiency Stress (NDS) (Figure 1). Once significant NDS has set in, the ability for the crop to recover and produce full yield is minimal. As a result, early detection of regions experiencing NDS is paramount to ensuring an optimal yield; better still is the ability to predict which regions may begin showing stress in the near future so they may be preemptively treated. On the other hand, overuse and blanket applications of fertilizer and other chemical nutrients can cause harm to the plants and additionally, runoff into the water table or other bodies of water, harming the environment. Therefore, early, precise and accurate identification of these regions is crucial for addressing both economic and environmental issues. We choose to focus on this task for the present work although the methods could easily be applied to other agronomic patterns such as weeds or low emergence.

Figure 2 shows the impact of NDS on a field's production and thus a farmer's yield. On June 2, a farmer sowed over 4.5million corn seeds across his 152 acre farm in the corn belt. Roughly a month later on July 3, he applied a combination of weed killer and fertilizer uniformly across the entire field (far left). By collecting high-resolution imagery (10cm/pixel) over the entire season, we are able to observe the changes in the field as it develops (middle plots). We see that many of the areas which experienced NDS during the season correspond to regions of lower yield (red) at harvest time in late November (far right); these red areas produced >195 bushels/acre (orange: 195-230 bu/ac) whereas the best areas (dark green) produced over 275 bushels/acre. Had the grower applied additional nutrient mid-season, some of this yield could have been recovered. However, as these areas are only a small portion of the field, blanket application would be costly as wasteful. Through targeted applications based on our models, the grower could instead apply fertilizers in a precise manner, minimizing cost, maximizing yield, and

[*] work done while at Intelinair, Inc.

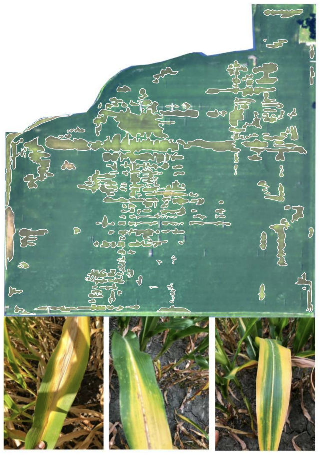

Figure 1: Nutrient Deficiency Stress appear brighter green in this aerial image which have been annotated with polygons(top). By alerting farmers to regions of the field experiencing NDS, further ground-level inspection can reveal what type of nutrient deficiency exists so the best management decisions can be made. Ground-level images of interveinal (bottom left), nitrogen (bottom middle), and potassium (bottom right) stress.

minimizing the runoff of chemicals into the water system.

To bring targeted intelligence to the grower in an actionable time frame, we collect high-resolution aerial imagery multiple times across the season. We first baseline our approach by building a simple U-Net-style segmentation model to detect areas of NDS. We then improve on this approach by leveraging the sequence of flights to create a spatiotemporal detection model. Finally, we show this same architecture can be used to predict areas of NDS in subsequent flights.

Related Work

Aerial Imagery, Remote Sensing, and Agriculture

Aerial Imagery and remote sensing techniques have been commonplace in the agriculture space for many years (Mulla 2013; Maes and Steppe 2019; Clevers 1986; Idso, Jackson, and Reginato 1977). Applications are widespread including high-throughput phenotyping (Araus and Cairns 2014), biomass prediction (Johansen et al. 2020), irrigation management (Bastiaanssen, Molden, and Makin 2000), weed detection (Thorp and Tian 2004), disease and pest detection (Zhang et al. 2019), nitrogen status and usage (Bausch and Duke 1996; Bausch and Diker 2001), and many others.

In applications related to agriculture, vegetative indices like the Normalized Difference Vegetative Index (NDVI) (Sharma et al. 2015) have been central to traditional computer vision based analyses (Xue and Su 2017; Huete et al. 2002). Another key index, Green Normalized Difference Vegetative Index (GNDVI) (Gitelson et al. 1996) is more strongly correlated with the concentration of chlorophyll and therefore to the rate of photosynthesis, making it a potential indicators of stress. The Normalized Difference Water Index (NDWI) (McFeeters 1996) is related to the water content in bodies, such as plants, and therefore can be informative about water-related stress.

As with many algorithms based on hand-crafted features, algorithms based on vegetative indices suffer from appearance changes due to variable lighting; in the imagery of Figure 2 seamlines are present, resulting in different broad appearances on the two sides of the image. This can cause challenges when relying on these indices unless other corrections and adjustments are made (Rodriguez et al. 2006; Noh et al. 2005; Mamaghani et al. 2018). We explore the use of indices for this deep learning task in *Experiments and Results: Vegetative Indices*.

Nutrient Deficiency Identification

Both traditional computer vision and deep learning methods have been used to detect and classify types of nutrient deficiency stress. Many of these results focus on identifying the type of deficiency from close-up images of the plant (Sartin, Da Silva, and Kappes 2014; Sethy et al. 2020). Early work on aerial imagery focused on identifying signatures in hyperspectral imagery and shortwave radiation correlated with the presence of NDS (Goel et al. 2003; Blackmer, Schepers, and Meyer 1995). Most of these cite changes in the 380-720nm or 720-1500nm range as strong indicators of stress due to changes in chlorophyll activity (Mee, Balasundram, and Hanif 2017). To the best of our knowledge, no work has been done to *forecast* NDS directly from aerial imagery.

Deep Learning in Agriculture

The adoption of deep learning methods for agricultural applications has accelerated in recent years (Kamilaris and Prenafeta-Boldú 2018; Liakos et al. 2018). These results can be largely split into those focused on "standard" imagery and those focused on aerial imagery from satellite, drone, or aircraft. Applications include disease and pest identification (Mohanty, Hughes, and Salathé 2016; Wiesner-Hanks et al. 2019; Boulent et al. 2019), crop identification (M Rustowicz et al. 2019), crop counting (Malambo et al. 2019; Li et al. 2017), weed detection (Sa et al. 2018; Bah, Hafiane, and Canals 2018; Sa et al. 2017), yield forecasting (Barbosa et al. 2020; Nevavuori, Narra, and Lipping 2019),, and parcel segmentation (Aung et al. 2020) as a few examples.

Specifically relevant to this work is (Chiu et al. 2020b) which used deep learning-based segmentation techniques to semantically segment the field into different patterns, including NDS, from high-resolution aerial imagery. Their approach used a DeepLabV3+ (Chen et al. 2018) model for their segmentation task. We similarly use an encoder-decoder structure, but focus on a U-Net (Ronneberger, Fischer, and Brox 2015) framework because of its success in other agricultural applications and computational efficiency (Lin and Guo 2020; Chiu et al. 2020a).

Spatiotemporal Modeling

Encoder-Decoder models like U-Net are commonplace in modern semantic segmentation tasks because of their performance, speed, and flexibility. More recently it has be-

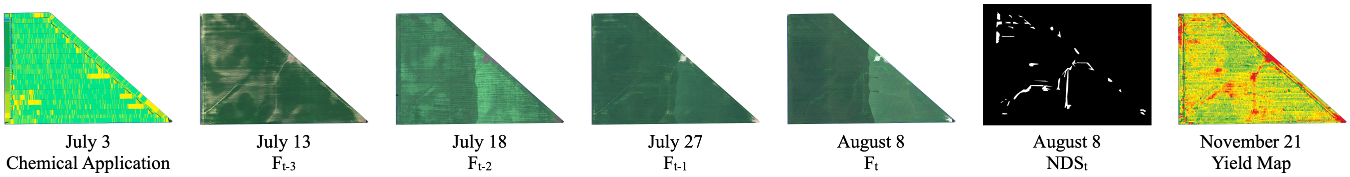

Figure 2: Temporal view of a field in this study. Corn was planted on June 2 and a uniform application of fertilizer and weed repellent applied on July 3 (far left). As the crop emerges and develops, under-performing areas due to nutrient deficiency stress (NDS) and other causes become visible during mid-season. Unless treated early, these nutrient deficient areas eventually under produce at harvest time(red, far right). Our model can be used to both detect the presence of NDS in a given flight F_t and predicts NDS in those areas from earlier flights $F_{t-3:t-1}$.

come common to use different backbones such as EfficientNet (Tan and Le 2019) within the U-Net framework. To address the temporal nature of data, Long Short Term Memories (LSTMs) (Hochreiter and Schmidhuber 1997) are frequently employed in a variety of deep learning-based sequence tasks including handwriting recognition, language translation, and action recognition, as a few examples (Graves et al. 2008; Wu et al. 2016).

Spatiotemporal modeling in the remote sensing domain is an active area of research (Zhu et al. 2017). Methods for handling the temporal element are highly varies and include: spatiotemporal U-Net (Lin Aung et al. 2020), histogram-based input representations (You et al. 2017), 3D convolutions (Ji et al. 2018), and many others. Our proposed method most closely resembles the Fully Convolutional Network(FCN)-LSTM network of (Teimouri, Dyrmann, and Jørgensen 2019), and was similarly chosen for its efficiency and performance.

Methods

Data Collection

Much of the work done in remote sensing for agriculture uses either low-resolution (10m/pixel) satellite imagery or very high-resolution (<5cm/pixel) imagery obtained from drones. While low-resolution satellite imagery provides a good overview of the field and scales across large areas, it does not provide enough resolution to precisely identify areas of the field which may exhibit stress. Conversely, while the imagery from drones is much higher, it is impractical to use to gather the desired data at such a low-resolution across millions of acres in a region.

Therefore to collect our data, fixed-wing aircraft were flown across corn and soybean fields in Illinois, Indiana, and Iowa, capturing imagery up to 13 times across the 2019 growing season (April to October). RGB and near-infrared (NIR) images were simultaneously captured at a resolution of 10cm/pixel using a Wide Area Multi-Spectral System (WAMS).

Mosaicking using ground-control points is performed to create a single large image per farm; depending on the farm size, this results in an image on average 15k-pixels × 15k-pixels in dimension. Orthorectification is performed using the RGBN image and a digital elevation model (DEM) of the field to produce a plainmetrically correct image (Gao, Masek, and Wolfe 2009). Geo-information is tagged to every image, but not used in this work nor a part of the data release to protect privacy.

Annotation and Dataset Construction

Images from 670 farm parcels were annotated for regions of nutrient deficiency stress by human experts; quality assurance (QA) of the annotations was conducted after. We focus only on mid-season flights, 6-10, when NDS is potentially present. 386 of the 670 flights contain at least 3 flights during this period and at least one flight demonstrating NDS; to conduct a fair comparison between single flight analysis and multiple flights analysis, we focus only on these 386 fields. The last annotated flight from each of the 386 fields becomes the target of subsequent analysis. Those 386 flights contained 10052 regions of NDS in total; not every flight showed signs of NDS while many had multiple regions. NDS is a relatively rare pattern spatially, resulting in an imbalanced dataset; on fields containing any NDS, an average 21% of those pixels contained NDS.

As the original images contain over 225 million pixels on average, we reduced the size by converting to 300dpi- high enough to preserve the pixel information and low enough to fit into our memory during training. This results in images roughly 1000-to-2000 pixels by 1000-to-2000 pixels with 1m/pixel resolution, still far higher than most satellite imagery.

As a part of this work we are releasing these three flights from the 386 farm parcels as well as the ground-truth mask for that final flight[1].

Data Augmentations: During training we perform data augmentation such as vertical/horizontal flipping, shifting, rotation, and padding using the Albumentations package (Buslaev et al. 2020). We do not perform any color-based augmentation because these images are narrow-band and therefore standard augmentation techniques do not produce the same results as in standard imagery. Note that in experiments where we use temporal data, we perform the same augmentation on all the images in the sequence.

[1] This dataset will be released on the Registry of Open Data on AWS under ”Longitudinal Nutrient Deficiency”. Supplementary material available at https://arxiv.org/abs/2012.09654.

Models

Single Timestep: To perform either NDS detection or prediction for only one time-point we use a U-Net structure with a VGG16 or EfficientNet backbone. Given the single input data point I, we have G as a ground truth mask and P as a predicted mask. The size of input I is w×h×c where c indicates the number of channels. Both the ground truth mask G and predicted mask P show the binary segmentation of NDS areas therefore they have a size of w×h×1. We calculate the combined Focal loss (Lin et al. 2017) + Dice loss (Sudre et al. 2017) for our final loss.

Multiple Timesteps: A key aspect of the collected data is its sequential nature which captures the evolution of the field over time. To incorporate the information from sequential flights, we construct the model seen in Figure 3. This model include 3 parallel U-Nets with EfficientNet backbones followed by a sequence of 2D convolutional-LSTM and Batch-normalization(BN) layers and a 3D Convolution layer to generate the final output. Given 3 consecutive flight images I_t, I_{t-1} and I_{t-2} and the final ground truth mask G_t belonging to time step t, each U-Net produces its respective binary mask S_t, S_{t-1} and S_{t-2}. These masks are stacked and passed through a convolutional-LSTM (many-to-many) layers. Finally the last 3D convolution layer generate the semantic segmentation mask using a sigmoid function, identifying NDS for each flight denoted by P_t, P_t^{-1} and P_t^{-2}. To force the output of each U-Net to resemble the final mask, not just to provide information to the next flight via the LSTM, we calculate the combined Focal + Dice loss for each of these predictions P_t, P_t^{-1}, P_t^{-2}. The total loss is used defined as:

$$\mathcal{L}_{Total} = \frac{1}{3}\sum_{i=0}^{2} Loss_{t-i}^{Focal} + Loss_{t-i}^{Dice}$$

Training: All experiments are conducted with a batch size of 2 for 200 epochs; all models reach their convergence point before the final epoch and the best model is chosen based on minimum validation loss. We use Adam optimization with an initial learning rate of 1e-4 and decayed it by a factor of 10 if no improvement to the validation loss was seen for a period of 10 epochs. We use Intersection Over Union (IOU) score and F1 score as two evaluation metrics to compare models. All the models are implemented using Keras (version 2.2.4) and Tensorflow (version 1.15) and we run them on 4 NVIDIA Tesla V100 GPUs with 64GiB memory in total.

Experiments and Results

Our experiments are as follows: first, we show different approaches to find the best model for single time step detection using latest flight and its NDS ground truth mask. In this evaluation, four scenarios have been considered including data augmentation, U-Net backbone, ImageNet pretraining and different vegetation indexes. In the next step, we show the results of our proposed model for multiple timesteps in both NDS detection and prediction along side with ablation studies results for comparison.

Single Timestep Baselines

We explore the impact of resolution on model performance by either rescaling or cropping the images. In the former, we rescale the full-field image to 512×512 using bilinear interpolation. Alternately, we crop the images to a fixed size of 512×512, thereby maintaining the full resolution. Cropping is performed in the training pipeline in one of two ways: "random" or "wise" crop. For random cropping, we select the 512×512 patch from full-size image randomly, resulting in many patches with no nutrient deficiency. In contrast, in our "wise crop" approach, we use a 512×512 patch for training only if it contains some NDS masks.

ImageNet (Deng et al. 2009) pretraining is common in computer vision applications because it can speed up training and may result in a better learned representation (Kornblith, Shlens, and Le 2019). However, since the statistics of remote-sensing imagery in general and aerial agricultural imagery in particular are dramatically different than ImageNet (Xie et al. 2015), we explicitly investigated its usefulness in pretraining in this domain. Therefore we conduct all the experiments here with and without pre-trained ImageNet weights and compared their performance on the test dataset.

For all the experiments we use U-Net framework and compare two different backbones: VGG16 (Simonyan and Zisserman 2014) and EfficientNet-B5 (Tan and Le 2019). We use data from the 386 flights randomly separated to: 231(60%) train, 77(20%) validation, and 78(%20) test. We used only the RGB channels, ignoring the NIR channel.

As seen in Table 1, the U-Net pretrained with ImageNet weights with an EfficientNet-B5 backbone using the "wise cropping" strategy produces the best results with an F1-score of 0.43, an IOU of 0.34, and a (Focal + Dice) loss of 0.58. While these results are slightly lower than those of (Chiu et al. 2020b), the two analyses are not directly comparable; the resolution and number of samples is lower in our analysis, we use only RGB instead of RGBN, and this is a single instead of multi-task approach. The goal of our baselining is not to outperform other single-timestep models, but to quantify the performance on this particular dataset and establish an understanding of what architectural and sampling strategies might best guide our longitudinal analysis.

Using pretrained ImageNet weights improves performance for each of the models. Although the statistics of ImageNet and our dataset are quite different as noted before, the ImageNet weights nevertheless lead to an improvement in performance; therefore all subsequent models are conducted using pretrained weights.

Comparing the impact of backbones, our analysis shows that the EfficientNet backbone continually outperforms the VGG16 backbone; therefore we focus on EfficientNet as the backbone to our U-Net in our longitudinal analysis.

Finally, these results show the importance of the wise crop sampling strategy over rescaling or random cropping. Regardless of the backbone and pretraining used, wise crop always lead to improved results; therefore we use this cropping strategy during our subsequent analyses. An example

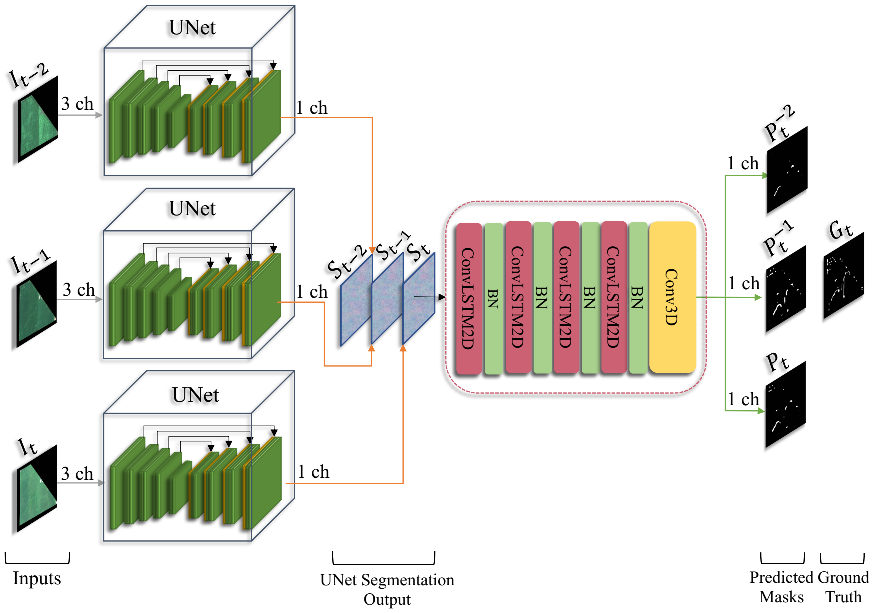

Figure 3: Our proposed model architecture for longitudinal detection of NDS. RGB images at three timesteps, I_t, I_{t-1}, I_{t-2}, are each passed through a U-Net to generate single channel outputs S_t, S_{t-1}, S_{t-2}. These outputs are stacked and then passed through a convolutional-LSTM layers which generates prediction P_t, P_t^{-1}, P_t^{-2} of NDS for flight t called G_t.

		F1 Score	**IOU Score**	**Loss (Focal + Dice)**
ImageNet Pretraining	VGG16 Full Rescaled	0.33	0.25	0.76
	VGG16 Random Crop	0.36	0.30	0.75
	VGG16 Wise Crop	0.38	0.30	0.71
	EfficientNet Full Rescaled	0.39	0.26	0.65
	EfficientNet Random Crop	0.40	**0.33**	0.64
	EfficientNet Wise Crop	**0.43**	**0.34**	**0.58**
No Pretraining	VGG16 Full Rescaled	0.28	0.19	0.80
	VGG16 Random Crop	0.30	0.26	0.76
	VGG16 Wise Crop	0.30	0.28	0.73
	EfficientNet Full Rescaled	0.31	0.21	0.74
	EfficientNet Random Crop	0.33	0.29	0.69
	EfficientNet Wise Crop	**0.36**	**0.29**	**0.68**

Table 1: Impact of backbone architecture, pretraining, and cropping on performance

of the output from this best model is shown in the top row of Figure 4.

Vegetative Indices

Input representation is central to successful deep learning models. Because vegetative indices play such a key role in traditional agricultural analysis and pattern detection, we next examine the impact and usefulness these indices have in a deep learning setting. Definitions of theses indices are provided in the Supplementary Material.

Given the results of the previous analysis, we use 512×512 images obtained with wise cropping, and a U-Net with an EfficientNet backbone, pretrained on ImageNet for all subsequent experiments. We compare four different input representations: RGB only (3 channel), NDVI (1 channel), RGB + NDVI (4 channel) with a 1D convolution applied at the first layer, and NDVI + GNDVI + NDWI (3 channel). Results are shown in Table 2.

These results show that using the raw RGB image as the input into the model far outperform any other input representation. NDVI alone produce the worst results. This is perhaps not surprising because it contains information from only 2 channels (Red and NIR) compare to the 3 RGB channels so there is necessarily less information present. However, given its prevalence in the remote-sensing domain and the reliance on this metric in past approaches, quantifying just how much

	F1 Score	IOU Score	Loss (Focal + Dice)
Image (RGB)	**0.43**	**0.34**	**0.58**
NDVI	0.25	0.15	0.85
Image(RGB)+NDVI	0.36	0.29	0.69
Image(RGB)+GNDVI	0.34	0.29	0.71
Image(RGB)+NDWI	0.33	0.26	0.75
NDVI, GNDVI, NDWI	0.32	0.24	0.73

Table 2: Impact of vegetative indices and image channels on performance

information is lost by using this single index is important.

Interestingly, using the combination of RGB and NDVI, which incorporates information from all four channels, with a 1D (channel-wise) convolution to create a new input representation for the U-Net, perform worse than RGB alone. We suspect this is due to losing too much information too quickly in this first 1D convolutional layer; that is, even by “learning” a new 1D representation as opposed to proscribing it through an index like NDVI, significant information is lost by quickly reducing the dimensionality.

Given the superior performance of the RGB representation, we focus on an RGB-only input for our longitudinal analysis.

Detection of NDS using Longitudinal Data

The best single timestep model from the earlier analysis is used to predict G_t from a single image. However this time after initializing the network using ImageNet pretraining, we freeze the layers of encoder to decrease network’s trainable parameters. This make single step analysis more comparable to the subsequent longitudinal study which are relatively larger and take more space from GPU’s memory to fit. Results are shown in Table 3. As expected, the performance of the detection task ($P_t : I_t \rightarrow G_t$) is comparable to the previous results. While information about G_t is contained in I_{t-1} and I_{t-2} as seen in the prediction tasks $P_t^{-1} : I_{t-1} \rightarrow G_t$ and $P_t^{-2} : I_{t-2} \rightarrow G_t$, respectively, the performance decreases substantially to a point that would be unusable for any real-world application. Note that the loss for these tasks corresponds to a loss for only one timestep whereas the other tasks in Table 3 capture the average loss from three timesteps.

Additionally, we stack all three flights to create a 9-channel image and again used the same U-Net framework. This model perform particularly poorly so we add a 1D Convolution after the input layer; this raise the performance to be only slightly better than the single image (I_t) model.

To incorporate the information from sequential flights, we use our proposed model seen in Figure 3 and examine the impact of shared vs unshared weights of the U-Nets. Additionally we compare this approach to three alternative approaches we call *Only-LSTM*, *Pre-LSTM* and *Cascading-model*. The Only-LSTM model contain only the LSTM part of our main proposed model. The Pre-LSTM model takes a raw input images in sequence directly into the convolutional LSTM then multiplies (Hadamarad product) or concatenates the results with original inputs and passes them to 3 parallel U-Nets. The Cascading-model with concatenation takes an image I_{t-2}, passes it through a U-Net to get a predicted mask, and then combine the predicted probabilities P_t^{-2} with the next image I_{t-1} by concatenating it as a 4^{th} channel. This is passed through a second U-Net to produce mask P_t^{-1}; this mask is concatenated with I_t, passes through another U-Net, and the final predicted mask P_t is produced. Weights are updated such that the loss from P_t is propagated through the entire network, those from P_t^{-1} are propagated only through the first two U-Nets, and those from P_t^{-2} are propagated only through the first U-Net. The Cascading-model using multiplication module is the same as the previous except that the concatenations are replaced by the Hadamarad product. The diagram of these models are shown in the Supplementary Material.

Our proposed approach outperform the alternative models across all metrics (Table 3). The predicted masks at each timestep are shown in Figure 4. The model with the shared weights slightly outperforms or matches the model with unshared weights on all metrics, but also has the advantage of being much smaller in size. Unsurprisingly, incorporating all three flights significantly outperforms any of the single step models. While the improvement is not surprising, the amount of improvement is: the IOU is almost double that of the single-step model.

One might expect that the current flight includes (almost) all necessary information for detection because it reflects the current status of the field. However, as discussed in section *Aerial Imagery, Remote Sensing, and Agriculture*, there are large changes to the global appearance of the field including natural development of the growing season as well as lighting and other noise effects. It is reasonable to believe that the sequence of images allows the model to better differentiate between features explaining changes due to the underlying NDS and these other sources of noise which impact larger regions of the field.

Prediction of NDS using Longitudinal Data

We next ask whether this architecture could be useful in *predicting* NDS in later flights. Using our same proposed architecture from the previous section, we train the model on images $I_{t-1}, I_{t-2}, I_{t-3}$ to predict the nutrient deficiency regions of P_t; we call this “Prediction $t^{1:3}$”. We go a step further and train another model on images $I_{t-2}, I_{t-3}, I_{t-4}$, again to predict the regions of P_t; we call this “Prediction $t^{2:4}$”. We also reference the single-step prediction tasks discussed earlier.

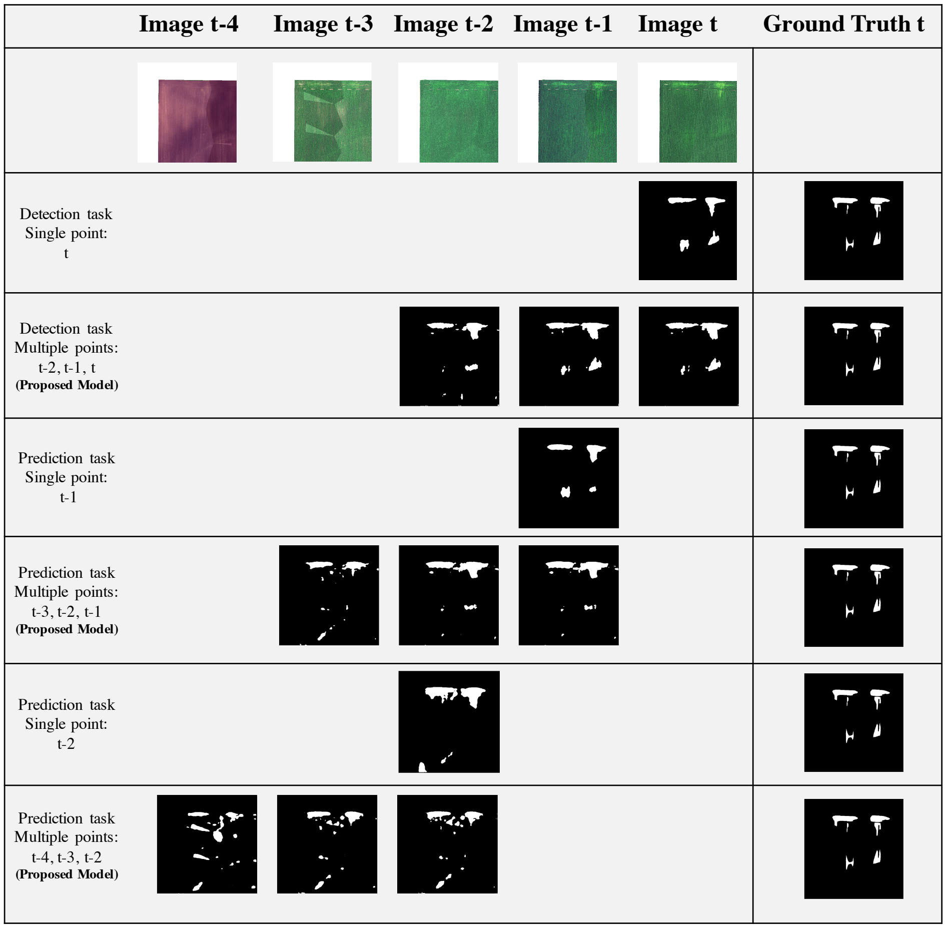
Image t-4
Image t-3
Image t-2
Image t-1
Image t
Ground Truth t
Detection task Single point: t
Detection task Multiple points: t-2, t-1, t (Proposed Model)
Prediction task Single point: t-1
Prediction task Multiple points: t-3, t-2, t-1 (Proposed Model)
Prediction task Single point: t-2
Prediction task Multiple points: t-4, t-3, t-2 (Proposed Model)

Figure 4: Results showing the predicted NDS mask P_t, P_t^{-1}, P_t^{-2} by using single flight model or sequential flights model (our proposed model).

		F1 Score	**IOU Score**	**Loss (Focal + Dice)**
Detection Task: I_t, I_{t-1}, I_{t-2}	Single ($P_t : I_t \rightarrow G_t$)	0.43	0.30	0.68
	9-Channel	0.23	0.15	0.90
	9-Channel + 1D Conv	0.48	0.34	0.62
	Proposed- Unshared	0.58	0.53	**0.85**
	Proposed- Shared	**0.62**	**0.57**	**0.85**
	Only-LSTM	0.48	0.43	0.90
	Pre-LSTM + Concat	0.42	0.29	0.92
	Pre-LSTM + Multi	0.42	0.29	0.89
	Cascading-Model + Concat	0.43	0.30	0.87
	Cascading-Model + Multi	0.38	0.25	0.91
Prediction Task: $I_{t-1}, I_{t-2}, I_{t-3}$	Single ($P_t^{-1} : I_{t-1} \rightarrow G_t$)	0.38	0.26	0.73
	Proposed- Unshared	0.55	0.52	0.91
	Proposed- Shared	**0.57**	**0.53**	**0.90**
Prediction Task: $I_{t-2}, I_{t-3}, I_{t-4}$	Single ($P_t^{-2} : I_{t-2} \rightarrow G_t$)	0.27	0.18	0.84
	Proposed- Unshared	**0.48**	**0.44**	**0.93**
	Proposed- Shared	0.46	0.43	**0.93**

Table 3: Performance of our longitudinal models for detection and prediction

Even on these prediction tasks the spatiotemporal model does quite well (Table 3, Figure 4). The loss for the the prediction task one flight out is 0.90 with an IOU of 0.53(shared) and the loss for two flights out is 0.93 with an IOU of 0.43(shared). As in the detection task, using shared vs. unshared weights does not show a significant difference, so we prefer the shared weights because of the reduced model size. Note that for this task of predicting two flights out, which can correspond to as much as 3 weeks into the future, the IOU of the predicted mask is better than even the best detection model made from a single image. This suggests that incorporating the temporal element of this data and allowing the model to learn how the field evolves over time has tremendous value even beyond providing the stability we saw in the longitudinal detection analysis.

Conclusion

This paper addresses the important task of identifying nutrient deficiency stress from longitudinal aerial imagery. The ability to not only detect, but *predict* regions of NDS in a field has tremendous economic value to the farmers as well as environmental impact and sustainability efforts. Our work shows that while a single image of the field is useful for detecting NDS, a sequence of images when used with an appropriate architecture provides significantly improved performance for both detection and prediction. While the present work has focused only on nutrient deficiency stress, we believe this framework will be useful for detecting and predicting other patterns of interest such as weeds, drydown, water, and others. Importantly, these models can be easily deployed at scale to a typical data pipeline for aerial agricultural imagery for maximum impact. They can easily be optimized for target hardware using frameworks which further improve inference speed and therefore reduce inference costs.

As this dataset is at a much higher resolution than most publicly available remote sensing imagery, we believe this will open the door to interesting future research. We only began to explore the usefulness of the NIR channel through our examination of vegetative indices; this and other studies around alternate or learned vegetative indices is the focus of ongoing work. While we explored a number of architectural variations, there is significant work being done on spatiotemporal data and sequential remote sensing imagery using completely different paradigms which this dataset will further enable. Our hope is that multiple research communities find this dataset useful in advancing both computer vision and sustainable agriculture.

References

Araus, J. L.; and Cairns, J. E. 2014. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in plant science* 19(1): 52–61.

Aung, H. L.; Uzkent, B.; Burke, M.; Lobell, D.; and Ermon, S. 2020. Farmland Parcel Delineation Using Spatio-temporal Convolutional Networks. *arXiv preprint arXiv:2004.05471* .

Bah, M. D.; Hafiane, A.; and Canals, R. 2018. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote sensing* 10(11): 1690.

Barbosa, A.; Trevisan, R.; Hovakimyan, N.; and Martin, N. F. 2020. Modeling yield response to crop management using convolutional neural networks. *Computers and Electronics in Agriculture* 170: 105197.

Bastiaanssen, W. G.; Molden, D. J.; and Makin, I. W. 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. *Agricultural water management* 46(2): 137–155.

Bausch, W. C.; and Diker, K. 2001. Innovative remote sensing techniques to increase nitrogen use efficiency of corn. *Communications in Soil Science and Plant Analysis* 32(7-8): 1371–1390.

Bausch, W. C.; and Duke, H. 1996. Remote sensing of plant nitrogen status in corn. *Transactions of the ASAE* 39(5): 1869–1875.

Blackmer, T.; Schepers, J.; and Meyer, G. 1995. Remote sensing to detect nitrogen deficiency in corn. In *Site-specific management for agricultural systems*, 505–512. Wiley Online Library.

Boulent, J.; Foucher, S.; Théau, J.; and St-Charles, P.-L. 2019. Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Frontiers in Plant Science* 10: 941. ISSN 1664-462X. doi:10.3389/fpls.2019.00941. URL https://www.frontiersin.org/article/10.3389/fpls.2019.00941.

Buslaev, A.; Iglovikov, V. I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; and Kalinin, A. A. 2020. Albumentations: Fast and Flexible Image Augmentations. *Information* 11(2). ISSN 2078-2489. doi:10.3390/info11020125. URL https://www.mdpi.com/2078-2489/11/2/125.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *ECCV*.

Chiu, M. T.; Xu, X.; Wang, K.; Hobbs, J.; Hovakimyan, N.; Huang, T. S.; Shi, H.; Wei, Y.; Huang, Z.; Schwing, A.; et al. 2020a. The 1st Agriculture-Vision Challenge: Methods and Results. *arXiv preprint arXiv:2004.09754* .

Chiu, M. T.; Xu, X.; Wei, Y.; Huang, Z.; Schwing, A. G.; Brunner, R.; Khachatrian, H.; Karapetyan, H.; Dozier, I.; Rose, G.; et al. 2020b. Agriculture-vision: A large aerial image database for agricultural pattern analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2828–2838.

Clevers, J. G. 1986. *Application of remote sensing to agricultural field trials*. Ph.D. thesis, Clevers.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

FAO, F.; et al. 2017. The future of food and agriculture–Trends and challenges. *Annual Report* .

Gao, F.; Masek, J. G.; and Wolfe, R. E. 2009. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing* 3(1): 033515.

Gitelson, A. A.; Kaufman, Y. J.; Merzlyak, M. N.; et al. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment* 58(3): 289–298.

Goel, P. K.; Prasher, S. O.; Landry, J.-A.; Patel, R. M.; Bonnell, R.; Viau, A. A.; and Miller, J. 2003. Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn. *Computers and electronics in agriculture* 38(2): 99–124.

Graves, A.; Liwicki, M.; Fernández, S.; Bertolami, R.; Bunke, H.; and Schmidhuber, J. 2008. A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence* 31(5): 855–868.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Huete, A.; Didan, K.; Miura, T.; Rodriguez, E. P.; Gao, X.; and Ferreira, L. G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment* 83(1-2): 195–213.

Idso, S. B.; Jackson, R. D.; and Reginato, R. J. 1977. Remote sensing for agricultural water management and crop yield prediction. *Agricultural Water Management* 1(4): 299–310.

Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; and Duan, Y. 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing* 10(1): 75.

Johansen, K.; Morton, M. J. L.; Malbeteau, Y.; Aragon, B.; Al-Mashharawi, S.; Ziliani, M. G.; Angel, Y.; Fiene, G.; Negrão, S.; Mousa, M. A. A.; Tester, M. A.; and McCabe, M. F. 2020. Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest. *Frontiers in Artificial Intelligence* 3: 28. ISSN 2624-8212. doi:10.3389/frai.2020.00028. URL https://www.frontiersin.org/article/10.3389/frai.2020.00028.

Kamilaris, A.; and Prenafeta-Boldú, F. X. 2018. Deep learning in agriculture: A survey. *Computers and electronics in agriculture* 147: 70–90.

Kornblith, S.; Shlens, J.; and Le, Q. V. 2019. Do better imagenet models transfer better? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2661–2671.

Li, W.; Fu, H.; Yu, L.; and Cracknell, A. 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing* 9(1): 22.

Liakos, K. G.; Busato, P.; Moshou, D.; Pearson, S.; and Bochtis, D. 2018. Machine learning in agriculture: A review. *Sensors* 18(8): 2674.

Lin, T.; Goyal, P.; Girshick, R. B.; He, K.; and Dollár, P. 2017. Focal Loss for Dense Object Detection. *CoRR* abs/1708.02002. URL http://arxiv.org/abs/1708.02002.

Lin, Z.; and Guo, W. 2020. Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning. *Frontiers in Plant Science* 11: 1346. ISSN 1664-462X. doi:10.3389/fpls.2020.534853. URL https://www.frontiersin.org/article/10.3389/fpls.2020.534853.

Lin Aung, H.; Uzkent, B.; Burke, M.; Lobell, D.; and Ermon, S. 2020. Farm Parcel Delineation Using Spatio-Temporal Convolutional Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 76–77.

M Rustowicz, R.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; and Lobell, D. 2019. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Maes, W. H.; and Steppe, K. 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends in plant science* 24(2): 152–164.

Malambo, L.; Popescu, S.; Ku, N.-W.; Rooney, W.; Zhou, T.; and Moore, S. 2019. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sensing* 11(24): 2939.

Mamaghani, B. G.; Sasaki, G. V.; Connal, R. J.; Kha, K.; Knappen, J. S.; Hartzell, R. A.; Marcellus, E. D.; Bauch, T. D.; Raqueño, N. G.; and Salvaggio, C. 2018. An initial exploration of vicarious and in-scene calibration techniques for small unmanned aircraft systems. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III*, volume 10664, 1066406. International Society for Optics and Photonics.

McFeeters, S. K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing* 17(7): 1425–1432.

Mee, C.; Balasundram, S.; and Hanif, A. 2017. Detecting and monitoring plant nutrient stress using remote sensing approaches: A review. *Asian J. Plant Sci* 16(1).

Mohanty, S. P.; Hughes, D. P.; and Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Frontiers in plant science* 7: 1419.

Mulla, D. J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering* 114(4): 358–371.

Nevavuori, P.; Narra, N.; and Lipping, T. 2019. Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture* 163: 104859.

Noh, H.; Zhang, Q.; Han, S.; Shin, B.; and Reum, D. 2005. Dynamic calibration and image segmentation methods for multispectral imaging crop nitrogen deficiency sensors. *Transactions of the ASAE* 48(1): 393–401.

Rodriguez, D.; Fitzgerald, G.; Belford, R.; and Christensen, L. 2006. Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts. *Australian Journal of Agricultural Research* 57(7): 781–789.

Rolnick, D.; Donti, P. L.; Kaack, L. H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A. S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. 2019. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433* .

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Sa, I.; Chen, Z.; Popović, M.; Khanna, R.; Liebisch, F.; Nieto, J.; and Siegwart, R. 2017. weednet: Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters* 3(1): 588–595.

Sa, I.; Popović, M.; Khanna, R.; Chen, Z.; Lottes, P.; Liebisch, F.; Nieto, J.; Stachniss, C.; Walter, A.; and Siegwart, R. 2018. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing* 10(9): 1423.

Sartin, M. A.; Da Silva, A. C.; and Kappes, C. 2014. Image segmentation with artificial neural network for nutrient deficiency in cotton crop. *Journal of Computer Science* 1084–1093.

Sethy, P. K.; Barpanda, N. K.; Rath, A. K.; and Behera, S. K. 2020. Nitrogen Deficiency Prediction of Rice Crop Based on Convolutional Neural Network. *Journal of Ambient Intelligence and Humanized Computing* .

Sharma, L. K.; Bu, H.; Denton, A.; and Franzen, D. W. 2015. Active-optical sensors using red NDVI compared to red edge NDVI for prediction of corn grain yield in North Dakota, USA. *Sensors* 15(11): 27832–27853.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* .

Sudre, C. H.; Li, W.; Vercauteren, T.; Ourselin, S.; and Cardoso, M. J. 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, 240–248. Springer.

Tan, M.; and Le, Q. V. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* .

Teimouri, N.; Dyrmann, M.; and Jørgensen, R. N. 2019. A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sensing* 11(8): 990.

Thorp, K.; and Tian, L. 2004. A review on remote sensing of weeds in agriculture. *Precision Agriculture* 5(5): 477–508.

Wiesner-Hanks, T.; Wu, H.; Stewart, E.; DeChant, C.; Kaczmar, N.; Lipson, H.; Gore, M. A.; and Nelson, R. J. 2019. Millimeter-level plant disease detection from aerial photographs via deep learning and crowdsourced data. *Frontiers in plant science* 10: 1550.

Wu, Y.; Schuster, M.; Chen, Z.; Le, Q. V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* .

Xie, M.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2015. Transfer learning from deep features for remote sensing and poverty mapping. *arXiv preprint arXiv:1510.00098* .

Xue, J.; and Su, B. 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors* .

You, J.; Li, X.; Low, M.; Lobell, D.; and Ermon, S. 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In *Thirty-First AAAI conference on artificial intelligence*.

Zhang, J.; Huang, Y.; Pu, R.; Gonzalez-Moreno, P.; Yuan, L.; Wu, K.; and Huang, W. 2019. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture* 165: 104943.

Zhu, X. X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; and Fraundorfer, F. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 5(4): 8–36.