HIGH SPATIAL RESOLUTION FOR CROP YIELD PREDICTION IN LARGE FARMING SYSTEMS: A NECESSITY OR ADDITIONAL OVERHEAD

Stella Ofori-Ampofo[1], Rıdvan Salih Kuzu[2], Xiao Xiang Zhu[1]

[1]Technical University of Munich
[2]German Aerospace Center (DLR)

ABSTRACT

Concerning county-level yield predictions in large farming systems, relying on coarse-resolution satellite images is customary. However, these images lack sufficient textural detail to accurately summarise spatial information. Our work evaluates the advantage of enhanced spatial resolution by conducting a comparative analysis between coarse-resolution, high-temporal-frequency MODIS data and relatively high-resolution, low-temporal-frequency Landsat data for predicting corn yield in the USA. We benchmark this comparison using several models in a spatial versus non-spatial input data case. According to our findings, incorporating high-spatial resolution in this context did not yield any significant benefits, as it comes at the cost of reduced temporal revisit. Root mean square error of about 13 bushels per acre can be achieved 2-4 weeks before harvest for less-intense drought years. Extreme drought-struck years are, however difficult to anticipate.

Index Terms— crop yield prediction, machine learning, convolutional neural network, recurrent neural network.

1. INTRODUCTION

Achieving zero hunger remains challenging due to climate extremes, economic shocks, and conflicts [1]. Therefore, precise and timely prediction of crop yields is crucial for decisions to control sudden food shortages. In this regard, open-access satellite data offers a wealth of spatio-temporal information that can be effectively utilized by integration of remote sensing and artificial intelligence. This combination has demonstrated great potential in various agricultural applications, such as crop condition assessment [2], vulnerability evaluation, and yield estimation [3, 4, 5].

There are studies utilizing different combinations of variables, such as satellite surface reflectance, vegetation indices, climate data [6, 7], as well as soil properties and farm management information [3], employing various machine learning (ML) and deep learning (DL) approaches. In large-scale farming systems, such as those found in the United States, coarse-resolution satellite images, often acquired from the Moderate Resolution Imaging Spectroradiometer (MODIS), are commonly utilized. To extract spatio-spectral and temporal features, volumetric convolutional neural networks (CNN) are typically employed [4, 7]. Alternatively, two-dimensional CNNs combined with recurrent neural networks (RNN) are used to model the temporal component, which generally outperforms individual CNN or RNN models [6, 8]. However, coarse-resolution images may lack sufficient textural information to capture representative spatio-spectral features. An advantage of higher spatial resolution is its ability to capture fine-grained variations, enabling the analysis of spatial patterns that can provide insights into crop textural variations. Their added benefit has been explored in similar tasks but for smaller farming systems, with shorter periods using synthetic data and without leveraging spatial insights [5, 9].

In this study, we present findings on county-level corn yield prediction by comparing coarse spatial yet high temporal resolution imagery (MODIS-MOD09A1.061) to higher spatial yet lower temporal resolution imagery (Landsat-8). The spatial resolution of MODIS and Landsat-8 is 500 and 30 meters with a temporal frequency of 16 and 8-day respectively. These surface reflectance products are complemented with weather variables and our baseline approach involves encoding without spatial information (pixel averages), and spatial order (time series as histograms) [4].

2. DATA

Corn production dominates in the mid-western US; hence, we limit our analysis to the top five production states: Iowa, Illinois, Indiana, Nebraska, and Minnesota. On average, farm sizes in these selected states are over 250 acres [10], which encourages using of coarse-resolution data for yield applications. Google Earth Engine [11] is used to collect surface reflectances and Daymet climate variables (precipitation, minimum, and maximum temperature) for the usual planting and harvesting period (April to September).

MODIS contains seven spectral reflectance bands within the visible, near infra-red and short-wave-infra-red regions. For Landsat, we retain six out of eleven bands to align with the spectral range of MODIS. The daily granularity of the Daymet variables are reduced to coincide with the 8-day and 16-day observation window of MODIS and Landsat. We fur-

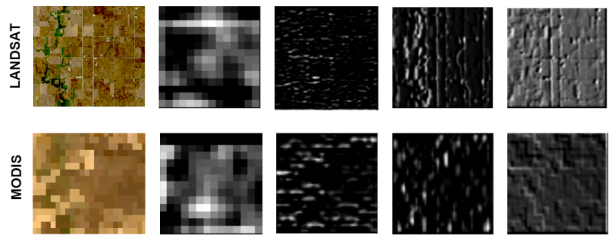

Fig. 1. Feature maps derived from convolution operations on Landsat and MODIS

ther compute normalized difference water (NDWI) and vegetation indices (NDVI) to introduce expert-knowledge features and mask non-corn pixels using crop data layer from the United States Department of Agriculture (USDA).

Two main data input formats are investigated: *(i)* a satellite time series $X \in R^{T \times C}$ created by averaging pixels within each county, *(ii)* a patch-level input of shape $X \in R^{T \times C \times H \times W}$. For patch inputs, we subsequently transform satellite image time series into compact three-dimensional histograms of binarized pixel counts $X \in R^{b \times T \times C}$ [4] where b is the number of bins and T, C, H, and W are the time series length, number of channels, image height and width respectively. The target variable, county-level average yield (in bushel per acre) is obtained from USDA, National Agricultural Statistics Service from period 2010 to 2021.

3. EXPERIMENTAL SETTING

We design a joint CNN-LSTM architecture (Figure 2) to simultaneously learn spatio-spectral and temporal patterns for patch inputs. The architecture consist of three blocks of time-distributed convolution layers, each followed by a batch normalization, rectified linear unit (ReLU) activation and max-pooling. The extracted spatio-spectral features are passed to an LSTM to learn temporal dependency. To avoid feeding the whole patch of a county (computationally challenging for Landsat data), five sub-patches are randomly sampled per county as inputs.

In the absence of spatial information, Random Forest (RF) is applied. Likewise, the CNN blocks shown in Figure 2 are removed to focus solely on modeling temporal sequences using LSTM. In the case of histogram inputs, two-dimensional convolutions are employed, following a similar approach as described in [4].

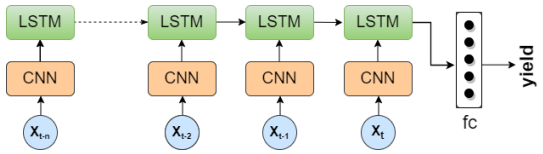

Fig. 2. Spatio-spectral and temporal feature extraction using a CNN-LSTM

Our experiments consist of two scenarios. *Scenario-1* considers only MODIS data from 2010 to 2021 across five states, increasing the overall training sample size. Scenario-1 includes observations from diverse agro-climatic zones and enables the comparison of different input data formats for test year 2020 and drought year 2012. [12]. Due to the absence of Landsat-8 data for 2012, *Scenario-2* strictly compares MODIS and Landsat data for Iowa state only. It considers observations from 2013 to 2021, with 2020 serving as an independent test set. The validation set comprises 25% of the training samples in all setups.

4. RESULTS AND DISCUSSION

4.1. Feature importance across time using Scenario-1

Using permutation-based importance on RF model (Figure 3), we observe for both cases of drought and non-drought the significance of NDWI during senescence. RMSE is reduced by an average of 4-5 bu/acre without the NDWI feature. The effect of the blue band, indicative of water presence, is seen to have a double impact during drought years. However, constructing a RF model using only the top significant features, NDWI and blue features, increases the RMSE to 44 and 25 bu/acre for 2012 and 2020.

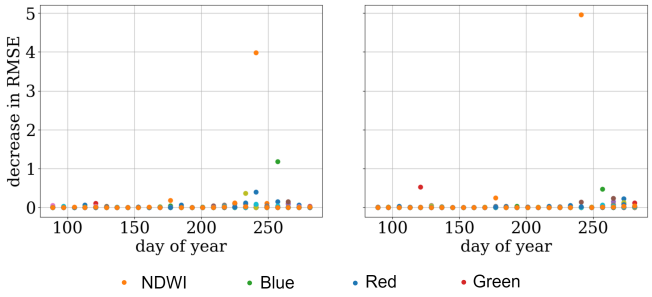

Fig. 3. Scenario-1: Temporal feature relevance using permutation importance on test sets. Drought-struck year 2012 (left), non-drought year 2020 (right). Legend is limited to the most significant features

4.2. Comparing performance for drought-struck and non-drought years

Drought year (2012) was more challenging to predict as it exhibits above-normal weather conditions during crop development, and presents an unprecedented sample in our time series. From Table 1, all models are competitive without extreme drought impacts but the LSTM explained over 50% of the variance for both test years. With extensive training samples (introducing geographical variability), we observe a general improvement in performance for the year 2020 compared to the same year in Scenario-2, where only Iowa state is considered. The latter generalises poorly to an unseen year. The relatively poor performance of CNN-LSTM in Scenario-1 can

Table 1. Performance metrics use only MODIS data (Scenario-1) over the top-five corn-growing states. Training data consists of satellite time series from 2010 to 2021 except for 2012 and 2020, which are independent test sets

Model	Validation			Test (2012)			Test (2020)		
	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2
RF (M)	$4.63^{\pm 0.04}$	$10.07^{\pm 0.06}$	$0.86^{\pm 0.01}$	$45.86^{\pm 1.1}$	$39.90^{\pm 0.91}$	$0.03^{\pm 0.05}$	$5.97^{\pm 0.15}$	$13.49^{\pm 0.27}$	$0.62^{\pm 0.02}$
Histogram-CNN (M)	$4.50^{\pm 0.0}$	$9.72^{\pm 0.23}$	$0.88^{\pm 0.01}$	$34.40^{\pm 0.01}$	$31.85^{\pm 0.68}$	$0.40^{\pm 0.03}$	$6.12^{\pm 0.0}$	$13.66^{\pm 0.25}$	$0.62^{\pm 0.01}$
LSTM (M)	$\mathbf{4.48^{\pm 0.0}}$	$\mathbf{9.86^{\pm 0.26}}$	$\mathbf{0.87^{\pm 0.01}}$	$\mathbf{30.70^{\pm 0.01}}$	$\mathbf{27.75^{\pm 0.79}}$	$\mathbf{0.53^{\pm 0.03}}$	$\mathbf{5.71^{\pm 0.0}}$	$\mathbf{12.78^{\pm 0.80}}$	$\mathbf{0.66^{\pm 0.04}}$
CNN-LSTM (M)	$5.78^{\pm 0.01}$	$12.84^{\pm 2.08}$	$0.79^{\pm 0.06}$	$55.08^{\pm 0.01}$	$50.31^{\pm 0.87}$	$-0.49^{\pm 0.05}$	$7.32^{\pm 0.0}$	$16.30^{\pm 0.08}$	$0.46^{\pm 0.01}$

Table 2. Comparison of Landsat (L) and MODIS (M) in Iowa state (Scenario-2). Training data consists of satellite time series from 2013 to 2021 except for 2020 which is independently used as a test set

Model	Validation			Test (2020)		
	MAPE	RMSE	R^2	MAPE	RMSE	R^2
RF (L)	$4.07^{\pm 0.17}$	$10.06^{\pm 0.38}$	$0.76^{\pm 0.02}$	$10.34^{\pm 0.30}$	$21.61^{\pm 0.47}$	$-0.82^{\pm 0.08}$
RF (M)	$3.45^{\pm 0.23}$	$8.04^{\pm 0.52}$	$0.86^{\pm 0.02}$	$11.98^{\pm 0.87}$	$24.24^{\pm 1.30}$	$-1.18^{\pm 0.23}$
Histogram-CNN (L)	$4.46^{\pm 0.0}$	$10.68^{\pm 0.05}$	$0.73^{\pm 0.01}$	$8.43^{\pm 0.01}$	$18.04^{\pm 0.85}$	$-0.30^{\pm 0.12}$
Histogram-CNN (M)	$4.19^{\pm 0.0}$	$9.95^{\pm 0.81}$	$0.78^{\pm 0.02}$	$7.73^{\pm 0.01}$	$16.66^{\pm 1.90}$	$-0.03^{\pm 0.23}$
LSTM (L)	$4.25^{\pm 0.0}$	$10.02^{\pm 0.2}$	$0.79^{\pm 0.01}$	$10.69^{\pm 0.01}$	$22.27^{\pm 1.42}$	$-0.94^{\pm 0.24}$
LSTM (M)	$\mathbf{3.30^{\pm 0.00}}$	$\mathbf{7.71^{\pm 0.53}}$	$\mathbf{0.87^{\pm 0.01}}$	$\mathbf{6.98^{\pm 0.01}}$	$\mathbf{15.36^{\pm 1.75}}$	$\mathbf{0.12^{\pm 0.21}}$
CNN-LSTM (L)	$9.50^{\pm 0.0}$	$21.14^{\pm 0.86}$	$0.01^{\pm 0.02}$	$10.4^{\pm 0.01}$	$21.53^{\pm 2.59}$	$-0.71^{\pm 0.41}$
CNN-LSTM (M)	$8.01^{\pm 0.0}$	$17.96^{\pm 0.7}$	$0.38^{\pm 0.01}$	$7.10^{\pm 0.0}$	$16.15^{\pm 0.19}$	$0.04^{\pm 0.02}$

be attributed to excluding a corn mask, leading to mixed land-use/cover in the patches and noisy signals. Applying a corn mask directly on the input creates spatial gaps due to non-contiguous corn farms.

4.3. Benefiting from improved spatial resolution

Using Landsat presents technical overhead, mainly in storage and computational requirements. For instance, a single-date Landsat observation for a county is 47 megabytes, approximately 280 times larger than MODIS. By comparing Landsat and MODIS (Scenario-2) under this setup, the highest performance is often seen with MODIS inputs except for the RF baseline. Similar to [9] there was no consistent value in using an improved resolution data for county-level corn yield prediction in large farming systems. Although Landsat offers a better level of spatial detail, it comes at the expense of reduced temporal resolution. Here, critical phenological states which may hold information necessary to estimate yield amount may be missed. Conversely, at farm-level, [13] deduced that using Landsat improved wheat prediction suggesting the appropriateness of higher spatial resolution for small spatial units.

4.4. In-season prediction

Tables 3 and 4 present the RMSE and R^2 metrics at the end of selected months using LSTM for Scenario-1. The year 2012 can be predicted well in advance (end of July), but four units can improve the performance for 2020 at the end of August. This improvement is achieved approximately 2-4 weeks before harvest, depending on the state.

Table 3. Comparing in-season RMSE using LSTM

	June	July	August	September
2012	62.61	27.80	28.79	**27.75**
2020	18.18	17.58	13.34	**12.78**

Table 4. Comparing in-season R^2 using LSTM

	June	July	August	September
2012	-1.40	**0.53**	0.49	**0.53**
2020	0.31	0.36	0.63	**0.66**

5. CONCLUSION

This study examined the benefits of improved spatial resolution for corn yield prediction by comparing MODIS and Landsat surface reflectance, complemented with weather variables. Our results suggest that spatially high, temporally low-resolution data offers no advantages for county-level yield assessments, while low-spatial, high temporal resolution data yields beneficial outcomes.

Regarding spatial-temporal feature extraction, the poor performance of the CNN-LSTM can be attributed to mixed or noisy land use in sampled patches, requiring an explanation layer to gain insights into the model's attention. In the case of Landsat, persistent cloud conditions can render an entire sequence of patches unusable. Such influence is minimal with