Deep Learning For Crop Yield Prediction in Africa

Apollo Kaneko* [1] **Thomas Kennedy*** [1] **Lantao Mei** [2] **Christina Sintek** [3] **Marshall Burke** [4] **Stefano Ermon** [1] **David Lobell** [4]

Abstract

Lack of food security persists in many regions around the world, especially Africa. Tracking and predicting crop yields is important for supporting humanitarian and economic development efforts. We use deep learning on satellite imagery to predict maize yields in six African countries at the district level. Our project is the first to attempt this kind of prediction in Africa. Model performance varies greatly between countries, predicting yields in the most recent years with average R^2 as high as 0.56. We also experiment with transfer learning and show that, in this data sparse setting, data from other countries can help improve prediction within countries.

1. Introduction

Improving food security and decreasing hunger by 2030 are key sustainability goals for the United Nations (Nations, 2018). However, the number of people without adequate food supply has been rising globally since 2015. In Africa, 22.7% of people are undernourished—significantly worse than the global average. The situation is especially dire in Sub-Saharan Africa, where it is estimated that 224 million people are undernourished (FAO, 2017).

Agricultural monitoring, and specifically the monitoring of crop yields, is vital in assessing food security in a region (Jones & Young, 2013). However, in developing countries, collecting high quality agricultural data can be expensive and difficult (Keita, 2019). In this paper, we address this problem by using publicly available remote sensing data to predict maize yields in six different African countries.

We use county level maize yield data from Ethiopia, Kenya, Malawi, Nigeria, Tanzania, and Zambia. Our remote data consists of satellite images from the MODIS collection (DAAC, 2018). To overcome sparse label data, we utilize a dimensionality reduction technique introduced by You et al. We convert raw images into histograms of pixel counts, which we then process using a LSTM-based deep learning model to predict crop yields. We also incorporate a Gaussian Process Layer to improve accuracy (You et al., 2017).

Our model performance varies by country. Testing on the last 4 years for Kenya, Tanzania, and Zambia, we achieve average R^2 values from .50 to .56. Our models perform worse on Ethiopia, Malawi, and Nigeria, with average R^2 values from -0.60 to 0.13. Combined models, where all countries are trained collectively, also show promising results with R^2 as high as 0.63 for randomized test sets. For some countries, the predictions from this model are more accurate than those made by individual in-country models. This suggests that models for countries with sparse training data can be improved by using out-of-country data.

2. Related Work

In early attempts to predict crop yields using remote sensing data, models relied solely on features such as normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) as inputs. Using this approach, Bolton and Friedl (2013) were able to predict maize yields in the United States with an R^2 of 0.69. More recently, deep learning models have been used to estimate yields. Kuwata and Shibasaki (2015) used a CNN with MODIS data to achieve an R^2 of 0.65 when predicting maize yields in Illinois. In 2017, You et al. implemented a novel dimensionality reduction technique, which we implement in this paper, to allow for model generated features. They also incorporated a Gaussian Processing Layer which further improved performance, and were able to predict soybean yields in the U.S. with RMSE ranging from 0.31 and 0.37 mt/ha (2017).

All of the aforementioned papers focus on the United States, where the ground truth yield data is reliable and easily ac-

*Equal contribution [1]Department of Computer Science, Stanford University [2]Department of Electrical Engineering, Stanford University [3]Stanford Center on Global Poverty and Development, Stanford University [4]Department of Earth System Science, Stanford University. Correspondence to: Apollo Kaneko <akkaneko@stanford.edu>, Thomas Kennedy <tkenned9@stanford.edu>, Lantao Mei <lantao@stanford.edu>, Christina Sintek <cssintek@stanford.edu>, Marshall Burke <mburke@stanford.edu>, Stefano Ermon <ermon@cs.stanford.edu>, David Lobell <dlobell@stanford.edu>.

Appearing at the International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States, 2019.

cessible. There has been some work done trying to predict yields in developing countries. Wang et al. (2018) trained deep learning models using MODIS data to predict soybean yields in Argentina and Brazil. They used the same dimensionality reduction technique as You et al. and achieved R^2 as high as 0.57 for Argentina and 0.66 for Brazil. We are not aware of any projects of this kind done in Africa.

3. Data

Our label data consists of yield data provided by the USAID Famine Early Warning System Network (FEWS NET) (Gary Eilerts, personal communication) at the Admin 2 (i.e. district) or Admin 1 (i.e. province) level for 6 countries. Yield is defined as metric tons of crop per hectare of land. Each label corresponds to a harvest season for a specific year. We remove data points with missing entries, yield outside of two standard deviations, and area planted less than 5000 hectares. As shown in Table 1, dataset size and yield distribution varies significantly by country.

Table 1. Yield distribution by country (mt/ha)

Yield Distributions by Country			
Country	Mean	Std. Deviation	n
Ethiopia	2.05	0.73	473
Kenya	1.52	0.95	532
Malawi	2.12	1.13	451
Nigeria	1.64	0.46	639
Tanzania	1.31	0.55	220
Zambia	1.77	0.89	1192

Our feature data consists of 500m resolution satellite images provided by the MODIS satellite collection. Specifically, we use all seven bands of the MOD09A1.006 Surface Reflectance data and two bands of the MOD11A2.006 Land Surface Temperature data representing day and night temperature (DAAC, 2018).

Because harvest seasons for maize differ between countries, we use the UN Food and Agriculture Organization country profiles to determine the length and time span for each country's growing season (FAO, 2018). We also validated the season period by plotting how the mean temperature and NDVI fluctuates over time.

There are some notable temporal trends in our label and feature data. We observe no significant trends between year and average peak NDVI. However, for Ethiopia, Malawi, and Zambia, there is a positive trend for yield over time. Other countries show no noticeable yield trend. These trends, shown in Figure 1, suggest that some temporal correction will help produce accurate predictions in certain countries.

(a) Yield Over Time (b) NDVI Over Time

Figure 1. Average NDVI and Yield Over Time for Ethiopia (orange) and Kenya (blue)

4. Approach

4.1. Problem Setting

We attempt to predict crop yields for a specific region over a given growing season and year. The feature space for each model is a collection of spectral images C. Each image I in C corresponds to a certain spectral band b and time step t. Therefore, for each label y, we have the feature set:

$$\left\{(I^1_{b_1}...I^t_{b_1})...(I^1_{b_n}...I^t_{b_n})\right\} \rightarrow y$$

We restrict images to be centered around the month with the peak NDVI over the course of the year. For our dataset, we take three months before the peak and two months after the peak to prioritize the growing period.

We train models for two kinds of splits: random and chronological. In the random splits, each pair (C, y) is randomly split into training, validation, and testing with the sets holding 80%, 10%, and 10% of the data respectively. In the chronological splits, we use data from the most recent year as the test set. We measure our performance by averaging over the prediction metrics for the most recent four years - each year trained on all preceding years. This split is more consistent with practical applications of this model, where one would train on existing data and predict future yields. Because many harvesting seasons have sparse labels, we train and test on the season with most data for each country.

4.2. Baselines

Our baseline model uses ridge regression, which is linear regression with L2 regularization. We use the surface reflectance and temperature data to generate features. From the surface reflectance data, we calculate NDVI and EVI values for each image over the course of the peak period. We do the same thing with the surface temperature images, except taking the mean pixel value for the images representing day and night temperature. We also test a simplified model with only two features: the highest average NDVI value and the average temperature during the growing period.

4.3. LSTM

4.3.1. Data Processing

Because the quantity of label data can be sparse, we use the histogram dimensionality reduction technique detailed in You et al. to combat overfitting. This technique assumes that the location of each pixel value within an image I is unimportant for this task relative to the quantity and distribution of pixel values. Let t represent the timestep for each image over a given period. Each image $I^{(t)}$ has dimensions $I^{h,w,d}$, where h and w are the dimensions and d is the number of bands. In our model, $d = 9$, with 7 surface reflectance bands and 2 surface temperature bands. We make the simplifying assumption that each band d is independent from the others. Then, for each d, we separate the pixel values into a histogram with b bins, where b is a hyperparameter. We use b = 32. Therefore, we represent each $I^{(t)}$ as a histogram $h^{(t)}$ with shape $b \times d$. We then stack each h for all t, resulting in a 3D histogram H of shape $T \times b \times d$, where T is the number of timesteps for a datapoint, b is the number of bins, and d is the number of bands (2017).

4.3.2. Model Structure

The full structure of our LSTM model is shown in Figure 2. For each histogram H, we flatten the feature in the bins and bands dimensions, preserving the temporal structure of the data. We feed the modified histograms into a LSTM network, applying dropout at each timestep. The outputs of the LSTM are then fed through two dense layers to output the yield prediction. We run a random hyperparameter search for each LSTM model. We use MSE loss, Adam Optimization, and Xavier initialization.

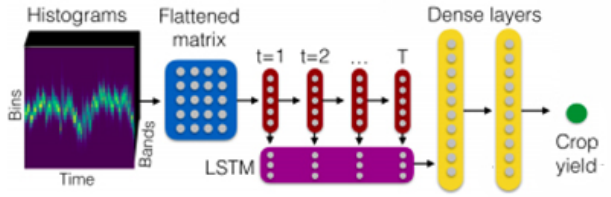


Figure 2. **LSTM Model Structure** (You et al., 2017). We feed a flattened 3D histogram into a LSTM network. The outputs are then fed into dense layers for prediction.

4.4. Gaussian Processing

There may be spatial and temporal trends in crop yield that are not captured in the satellite features. We integrate spatio-temporal features into our models using a linear Gaussian Process, as done in You et. al (2017), to model these errors.

A Gaussian Process is a probabilistic model defined as a set of random variables, $f(x)_{x \in X}$, where a finite subset has the joint Gaussian distribution

$$f(x) \sim GP(E[f(x)], cov(f(x), f'(x)))$$

For our paper, we use the spatio-temporal data (coordinates, years) for our covariance function, and use a variant of GP that is linear in respect to the features extracted from our deep model. The linear model is defined as

$$y(x) = f(x) + h(x)^T \beta$$

where $f(x) \sim GP(0, k(x, x'))$. The kernel function, $k(x, x')$, is created using the spatial temporal features and a squared exponential kernel, represented as

$$\sigma^2 \, exp\left[-\frac{\|g_{loc} - g'_{loc'}\|_2^2}{2r_{loc}^2} - \frac{\|g_{year} - g'_{year'}\|_2^2}{2r_{year}^2} \right] + \sigma_e^2 \delta_{g \cdot g'}$$

The trailing term represents additional Gaussian noise, defined as a product of the variance σ_e^2 and the Kroennecker delta. $h(\cdot)$ is a set of basis functions from the last dense layer of our deep model. Finally, $\beta \sim \mathcal{N}(w, \sigma_b I)$, where w are the trained weights from the final dense layer in our model. $\sigma, \sigma_b, \sigma_e, r_{loc}, r_{year}$ are hyperparameters, and we train using log marginal likelihood.

5. Results and Discussion

5.1. In-Country Models

In-Country models are trained and tested on individual countries. We find a large amount of variability in the performance of these models depending on the country. Our results across models are presented in Tables 2 and 3. The LSTM with Gaussian Processing consistently outperforms our other models. This suggests that there is large variations between yields over spatial and temporal dimensions.

5.1.1. Random Splits

The performance of models trained and tested on randomized splits are shown in Table 2. The LSTM with Gaussian Processing performs the best across metrics. These models generally perform better than models trained on chronological splits. However, randomized training and testing is not a realistic problem. In a real world setting, models will typically predict yields for the most recent year, for which there is no training data available.

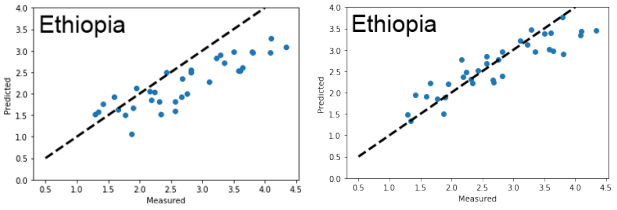
5.1.2. Chronological Splits

While the LSTM and Baseline do not perform well overall on Chronological splits, the Gaussian Processing improves the results. The full results are displayed in Table 3. Our models without any temporal features do poorly on countries like Ethiopia, where there is a trend over time not captured in the satellite data. Our frequent negative R^2 values for

Table 2. Results by Country for Random Splits

R^2				
Country	Simple	Ridge	LSTM	LSTM+GP
Ethiopia	0.09	0.49	0.48	**0.74**
Kenya	0.34	0.62	0.72	**0.75**
Malawi	-0.01	-0.01	0.11	**0.41**
Nigeria	0.12	0.11	0.34	**0.74**
Tanzania	-0.16	-0.01	0.24	**0.48**
Zambia	0.22	0.42	0.55	**0.62**
Pearson's r				
Country	Simple	Ridge	LSTM	LSTM+GP
Ethiopia	0.39	0.75	0.76	**0.86**
Kenya	0.59	0.80	0.86	**0.87**
Malawi	0.34	0.37	0.46	**0.66**
Nigeria	0.41	0.40	0.60	**0.87**
Tanzania	0.38	0.53	0.61	**0.70**
Zambia	0.52	0.68	0.76	**0.80**

these models suggest overfitting. Figure 3 shows that while the model with only satellite features is under-predicting due to the increasing yield over time, the model with Gaussian Processing is able to account for this trend.



(a) Without GP (b) With GP

Figure 3. **LSTM Performance on Ethiopia With and Without Gaussian Processing**: Plots of yield vs prediction

Our models perform significantly worse on Malawi and Nigeria. There are a few possible causes for this. Malawi has a relatively small dataset, which results in larger variation and noise between the training and testing data. Nigeria, on the other hand, has relatively small variance between yields, which results in low RMSE values but also low correlation scores. As shown in Figure 7, the Nigeria predictions are concentrated while the Malawi predictions are more widespread. This reflects how the model performance is affected by the underlying distribution of data. There is also the question of data quality. A recent study of the Malawi Food Subsidy Programme showed that maize yield numbers provided by the government are likely inaccurate (Messina et al., 2017). This would explain our consistently poor performance on Malawi across all models.

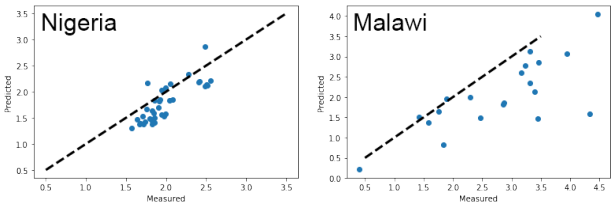


Figure 4. **LSTM+GP Performance on Nigeria and Malawi**: Plots of yield vs prediction

Table 3. Results by Country for Chronological Splits

R^2				
Country	Simple	Ridge	LSTM	LSTM+GP
Ethiopia	-1.41	-0.52	-0.35	**0.13**
Kenya	0.13	0.40	0.49	**0.56**
Malawi	**-0.03**	-0.54	-0.29	-0.09
Nigeria	-1.45	-3.44	-0.68	**-0.60**
Tanzania	0.03	-1.63	0.40	**0.50**
Zambia	0.23	-0.08	0.39	**0.56**
Pearson's r				
Country	Simple	Ridge	LSTM	LSTM+GP
Ethiopia	0.69	0.64	0.74	**0.82**
Kenya	0.42	0.69	**0.82**	**0.82**
Malawi	0.11	0.10	0.25	**0.55**
Nigeria	0.26	-0.03	0.40	**0.53**
Tanzania	0.38	0.11	0.74	**0.80**
Zambia	0.75	0.55	0.67	**0.77**

5.2. Transfer Learning

We also train our LSTM model on data from all six countries to see how features learned from one country can help predict yield for another. We use the concatenated training data of the random splits for training, and then predict upon the test sets of individual countries. These results are presented in Table 4. We opted not to use GP because we wanted to test the transferrability of satellite features without the influence of one country's spatio-temporal features. The full model achieves around 0.5 R^2 and 0.75 r when predicted upon the combined test sets. Three countries achieve higher metrics when trained on the combined model, with only Tanzania showing a notable drop in performance, which suggests that out-of-country features are improving the prediction.

6. Conclusion

We are able to predict yields in African countries with mixed success. Over all countries and tests, the LSTM with Gaussian Processing produces the best results. Using randomized splits, all models achieve high levels of accuracy. With the chronological splits, performance varies, possibly due to differences in feature and label distribution and data quality. Our combined model shows that a collective model

Table 4. In-Country vs Combined Model for Random Splits.

Country	In-Country		Combined	
	R^2	r	R^2	r
Ethiopia	0.48	0.76	**0.63**	**0.80**
Kenya	**0.72**	**0.86**	0.68	0.84
Malawi	0.11	0.46	**0.36**	**0.60**
Nigeria	0.34	0.60	**0.58**	**0.77**
Tanzania	**0.24**	**0.61**	-0.96	0.42
Zambia	**0.55**	**0.76**	0.49	0.71

performs competitively with in-country models, suggesting that the model can learn important out-of-country features.

Overall our models show that it is possible to predict crop yields using these methods at a relatively high level of accuracy. The results from Gaussian Processing model also emphasizes the importance of incorporating temporal spatial features when predicting crop yields.

For future study, different architectures such as a CNN model run on the histograms or the raw images may also attain significant results.

References

Bolton, D. K. and Friedl, M. A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173:74–84, 2013.

DAAC, N. L. Modis, 2018.

Damianou, A. C. and Lawrence, N. D. Deep gaussian processes. 2013.

FAO. Africa regional overview of food security and nutrition report. 2017.

FAO. Country profiles. http://www.fao.org/countryprofiles/en/, 2018.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017. doi: 10.1016/j.rse.2017.06.031. URL https://doi.org/10.1016/j.rse.2017.06.031.

Hochreiter, S. and Schmidhuber, J. Long short term memory. *Neural Computation*, 9(8):17351780, year = 1997.

Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

Jones, N. P. and Young. What are we assessing when we measure food security? a compendium and review of current metrics. *Adv Nutr. 4(5): 481505.*, 2013.

Kaneko, K. and Mei. Africa yield. https://github.com/akkaneko/africa-yield, 2018.

Keita, N. Improving cost-effectiveness and relevance of agricultural censuses in africa: Linking population and agricultural censuses, 2018.

Keita, N. Improving cost-effectiveness and relevance of agricultural censuses in africa: Linking population and agricultural censuses. 04 2019.

Kuwata, K. and Shibasaki, R. Estimating crop yields with deep learning and remotely sensed data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pp. 858–861. IEEE, 2015.

Messina, J. P., Peter, B. G., and Snapp, S. S. Re-evaluating the malawian farm input subsidy programme. *Nature plants*, 3(4):17013, 2017.

Nations, U. Transforming our world: the 2030 agenda for sustainable development. new york: United nations., 2018.

Pryzant, R., Ermon, S., and Lobell, D. B. Monitoring ethiopian wheat fungus with satellite imagery and deep feature learning. In *CVPR Workshops*, pp. 1524–1532, 2017.

Shastry, A., Sanjay, H., and Bhanusree, E. Prediction of crop yield using regression techniques. *International Journal of Soft Computing*, 12(2):96–102, 2017.

Wang, A. X., Tran, C., Desai, N., Lobell, D., and Ermon, S. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 50. ACM, 2018.

You, J. crop yield prediction. GitHub, 2017.

You, J., Li, X., Low, M., Lobell, D., and Ermon, S. Deep gaussian process for crop yield prediction based on remote sensing data. In *AAAI*, pp. 4559–4566, 2017.