CS230

Crop Yield Estimation of Rice for India
A Satellite-Based Deep Learning Approach

Ifeoma Anyansi, Adeesh Goel, and Agneya Loya
Institute for Computational and Mathematical Engineering
Department of Management Science and Engineering
Department of Civil and Environmental Engineering
Stanford University
`(ianyansi,adeesh,agneya)@stanford.edu`

Abstract

Crop yield forecasting plays a vital role in the pursuit of sustainable development. Predicting crop yields, such as wheat or rice, can help municipal governments plan out food sources and distribution to its population, and can serve as an early indicator of famine. In this paper, we apply deep learning models to predict rice yields in India; this represents a different development context from previous applications that focus on the United States. By using nine bands of multi-spectral satellite images and land surface temperature collected every eight days, we are able to achieve promising results given the lower granularity of our data. We find that both Convolutional Neural Networks and Long-Short Term Memory Networks are able to achieve high prediction accuracies with comparable performances. In contrast to previous research, we find that shallower models performed better, which may indicate that geographies with uncertain and sparse data may benefit from less complex model architectures to achieve stronger performance.

1 Introduction

Crop yield forecasting plays a crucial role in promoting sustainability and food security on a global scale. With the singular complexities and intersecting networks involved in food systems, forecasting activities can significantly reduce associated risks in supply, demand, transport, and storage [3]. With accurate and frequent food yield predictions, governmental entities and supply-chain actors are able to make more informed decisions that can increase food production and accessibility. The United Nations has emphasized the importance of this prediction task in one of their sustainable development goals to end hunger, achieve food security, and promote sustainable agriculture [4]. With the ever-growing population and the threat of climate change on sensitive agricultural systems, the advancement of reliable yield prediction methods is particularly significant in developing countries where implementation of modern forecasting techniques are too expensive and localized agricultural productivity is directly related with the health and lifestyle of surrounding communities.

With widespread and publicly available remote-sensing data, deep learning approaches can leverage unstructured image data from satellites to forecast crop yield. Multi-spectral satellite imagery contain both visible and non-visible wavelength of light including infrared and thermal. It has been shown that healthy vegetations reflect different wavelengths, thus satellite data is able to capture these differences while deep learning models are able to detect useful features to perform yield prediction [5]. Due to the quick and inexpensive nature of deep learning approaches and public satellite data, predicting crop yield with high accuracy can have immediate, far-reaching impact.

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

In this paper, we benefit from recent research in the field that uses satellite-based deep learning methods to forecast crop yield in the United States []. These techniques trained with bigger neural networks were shown to outperform traditional methods. We plan to use this existing framework as a starting point to predict rice yields in India by training two different neural network architectures of Convolutional Neural Networks (CNNs) and Long-short Term Memory Networks (LSTMs) on dimensionally-reduced satellite images. India presents a unique opportunity to evaluate the performance of deep learning methods in geographies where frequent and precise data concerning crop yield, agriculture, land, soil and temperature may not be readily available. In this paper, we find that rice yield data can be forecasted with relatively high accuracy which validates the use of these techniques in future developmental contexts.

2 Related work

Crop yield prediction has been previously performed using a diverse range of methods. Successful attempts at applying deep learning for this task have been a fairly recent development over the past year and hold great promise for the field.

2.1 Traditional approaches for crop yield forecasting

Crop yield prediction through remote sensing has been investigated since the late 1980s. The approach most investigated in that period was running linear regression models on hand-crafted features of remote sensing data. Quarmby et al. achieved a high degree of accuracy in predicting the yield for wheat, cotton, rice and maize crops using a linear regression between a feature called normalized difference vegetation index (NVDI) and yield [7]. The NVDI can detect the presence of live green vegetation in an image. Similarly, Prasad et al. used NDVI, soil moisture, surface temperature and rainfall data for Iowa to predict yield for corn and soyabean using a liner regression model [6].

2.2 Deep learning approaches for crop yield forecasting

Building on these traditional feature engineering based approaches, the past year has seen the emergence of end-to-end deep learning models using raw remote sensing data. You et. al, (2017) developed a successful deep learning framework for predicting soyabean yield that outperformed previous models that were based on hand-crafted features. They achieved high prediction accuracies by using CNNs along with a Deep Gaussian Process approach. This work was taken further by Sabini, Rusak and Ross (2017) where they implemented a crop differentiation algorithm and achieved lower error percentages by training deeper models. The success of these research groups further motivated us to implement their deep learning framework to yield prediction for India.

3 Dataset and Features

Multi-spectral satellite images for the Indian region were collected from NASA's MODIS instrument facilitated through the use of Google Earth Engine. MODIS images are available globally at a spatial resolution of 500m and are collected every eight days which sum to 46 times per year. Each image contained information for seven spectral bands, in which each band is a separate image that captures different wavelength ranges of light – from the visible to non-visible. As a reference, a common multi-spectral image is the RGB image where a particular image is represented by it's red, green, and blue bands respectively. Secondly, in order to include more predictive features in our model, we obtained satellite images for two additional bands of daytime and nighttime land surface temperature. Lastly, to filter these satellite images to only contain relevant data, that is satellite images solely corresponding to farmland in India, we retrieved yearly land cover masks from the MODIS satellite. By applying the mask on these images, we were able to extract 9-band images consisting of farmland throughout India. We further filtered our dataset to 32 of the 46 timesteps per year, to include images from February to October, thus capturing the Kharif rice cropping season and preceding months.

The crop that we chose to forecast district-specific yields for is rice. Not only is rice one of the most staple crops worldwide, it was the crop that possessed the largest share of historical yield information in India. Our ground truth data for historical crop yields was collected from a public database maintained by the Indian Government. The database provides historical yield data (tons/hectare) for

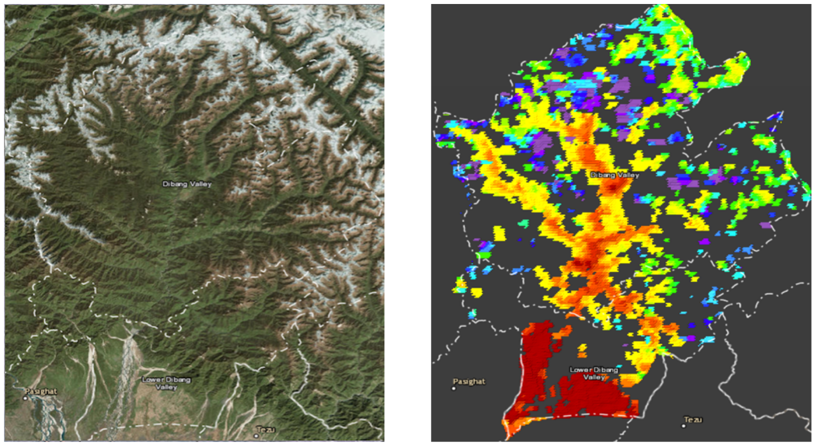

Figure 1: Dibang Valley in Arunachal Pradesh State (left: RGB bands, right: Temperature bands)

state-district-year combinations for all major crops since 1997. Although seasonal information were recorded, sufficient cleaning needed to be performed in order to standardize the data for use in our model. For example, since seasonal data wasn't consistent – some observations only reported yield numbers for certain seasons and not others – we opted to only consider the Kharif harvesting season in India which lasts from June to October. Rice is mainly grown in the Kharif season, and thus we were able to filter our locations and refine our prediction task accordingly to predict yield for this season. After all cleaning, we were left with 211 districts for which we collected raw satellite images from the years 2003-2009. The end result of this was 1323 data points.

If we were to use the raw images as the input, we would face significant dimensionality challenges. Since the satellite imagery across districts are large, distinct and include multiple time steps and bands, the total amount of features inputted to our model for just one training example would be on the orders of millions [8]. Thus a data transformation would first need to be performed to sufficiently reduce the dimensions of our input without a significant loss of the predictive properties found in the raw images.

An innovative technique for dimensionality reduction of problems of this kind was introduced in You et. al [9]. The key assumption is that of permutation invariance. This assumption hypothesizes that the location of pixels does not matter in predicting crop yields, solely the count of pixels do. In simpler terms, it is the aggregate number of green pixels in your image that is relevant in predicting crop yields versus the position of said pixels. Although this assumption may not integrate marginal localized properties that may exist across locations, we expect negligible loss of predictive power when evoking this assumption. Thus, we followed previous research and binned the raw images into a histogram where pixel value frequencies were encoded into 32 bins. 9-band images for each time step and location-year combination were transformed into a histogram array of shape 32 x 9 (i.e. # of bins x # of bands) in which rows quantified the number of pixels in 8-value ranges. The final input into our model was a 32x32x9 array which represented the number of bins, time steps, and bands respectively for a specific district-year observation. Our final output to our model were rice yield values corresponding to district-year observations.

4 Methods

In the realm of available deep learning models, we chose to focus our efforts on CNNs with our 3x3 filters convolving over the bins and time steps of our input shape (bins, time steps, bands) since CNNs resulted in the best performance for You et al. [9]. Independent of this approach we also implemented LSTM networks to account for the temporal component of our images and found them to perform better on outliers but equally well when averaged for the validation set. We found that given the scarcity of data, our simpler models tended to outperform more complex models, and so we focused our hyperparameter tuning on those.

4.1 CNN Models

Our best model included 2 convolutional layers, 1 max pooling layer, and 1 fully connected layer. Each CONV(c,f,s) layer represents a convolutional layer that has c filters of size f x f and a stride of s. This is followed by a ReLU non-linearity. The MAXPOOL(2,1) layer represents a max pooling layer of d 2 x 2, and a stride of 1. Finally, a fully connected layer with size 512 was used.

We explored two key error metrics, root mean squared error (RMSE) and mean absolute percentage error (MAPE). Although MAPE is more interpretable, it is also more sensitive to outliers and yields below one. Hence, we chose RMSE as our key error metric for the validation set, but consistently tracked MAPE for further interpretability.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(pred_i - real_i)^2} \quad | \quad MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{real_i - pred_i}{real_i}\right| \tag{1}$$

CNN Model Architectures			
Layer	Best Model	Simpler Model	Deeper Model
CONV(64,3,1)	1	1	1
MAXPOOL(2,1)	0	1	1
CONV(128,3,1)	1	0	1
MAXPOOL(2,1)	1	0	1
CONV(256,3,1)	0	0	1
FC(512)	1	1	1

Figure 2: Model Architectures for CNNs

4.2 LSTM Models:

The LSTM model used here is a many-to-one model that takes the 32 images spanning 256 days as a sequential input and outputs a single yield value. The typical architecture for the models consisted of initial many-to-many LSTM layers (denoted by *), feeding into a single many-to-one LSTM and terminating in fully connected layers. We conducted a manual grid search over the learning rate, batch size, L2 regularization, and weight decay hyperparameters.

LSTM Model Architectures			
Layer	Best Model	Deep 1	Deep 2
LSTM(32)*	2	2	2
LSTM(64)*	0	1	2
LSTM(64)	0	1	0
LSTM(32)	1	0	1
Dense(32)	0	1	0
Dense(10)	1	0	1
Dense(1)	1	1	1

Figure 3: Model Architectures for LSTMs

5 Experiments/Results/Discussion

After initial efforts of running deeper models with more than 6 layers, we learned that simpler models performed significantly better given the small size of the dataset. Thus, we focused our hyperparameter tuning on small models using dropout, L1/L2 regularization, and early stopping in some cases. The Adam optimizer performed best and after testing various batch sizes we found that the batch size of 25 gave us quickest convergence.

5.1 CNN Model

We achieved a best RMSE of 0.42 tons/hectare, and an MAPE of 33 percent. We can see in the loss chart below that the validation loss decreased steeply during the initial epochs and then started to stabilize (eventually underwent early stopping). The predicted versus actual chart (validation) shows the model's predictive capability, however, it does seem to perform relatively poorly on outlier values. The mean absolute percentage error declines consistently.

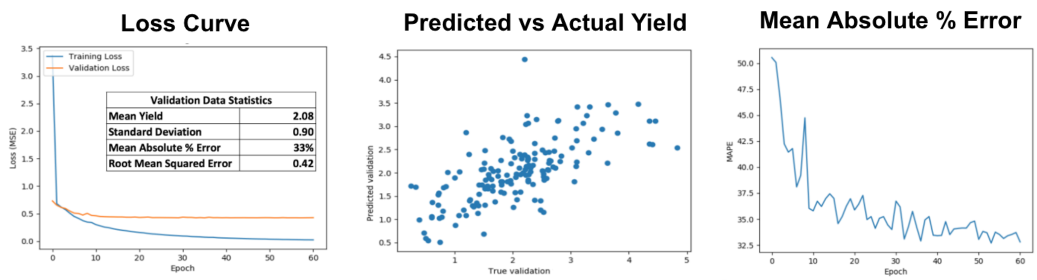

Figure 4: Best CNN Model Results

5.2 LSTM Model

We achieved a best RMSE of 0.44 tons / hectare, and an MAPE of 32 percent. Given that our data is sequence data (images over 8 day timesteps), it made sense that the LSTM model would work well. While the performance is overall quite similar to the CNN model, we do see that the validation loss starts to come down sooner, and the level of overfitting is lower. Moreover, the LSTM appears to perform better on extreme values as we can see in the chart below that very high and low yield values are closer to the diagonal axis of the predicted versus true yield chart.

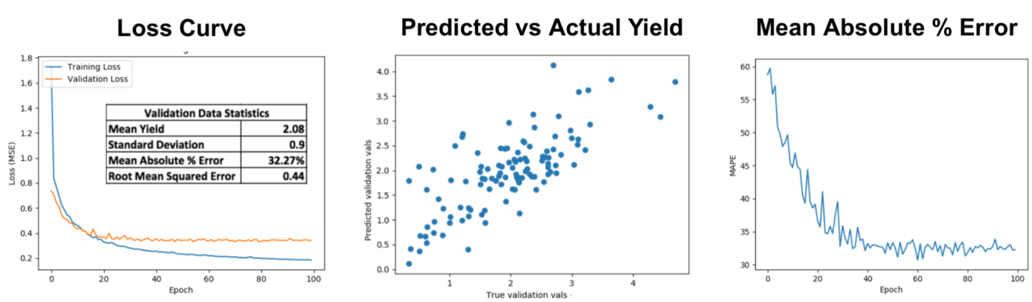

Figure 5: Best LSTM Model Results

5.3 Comparison of CNN and LSTM Performance on Extreme Yield Values

While the overall MAPE is close to 33 percent, this number is significantly influenced by outlier values and especially very low values of yield below one that inflate the mean absolute percent error. If we ignored outlier values and yields below one, we would find that the overall MAPE would be closer to 20 percent. This phenomena is due to the fact that the MAPE error metric heavily distorts percentage errors when actual values are near zero: often leading to undefined or infinite MAPEs [1]. Since our data is reported in tons/hectare which is standard in India, most of our ground truth yield values are on on the lower range and thus is heavily impacted by this issue. Secondly, although the CNN and LSTM models had comparable overall performance in terms of RMSE and MAPE, we find that the LSTM performs better on extreme values. Figure 6 shows the LSTM model is able to predict a higher proportion of yield values less than 1 with its percent error being less than 40 percent. Since farmers and government officials are most concerned with predicting extreme conditions such as famine, this model may potentially be more suitable.

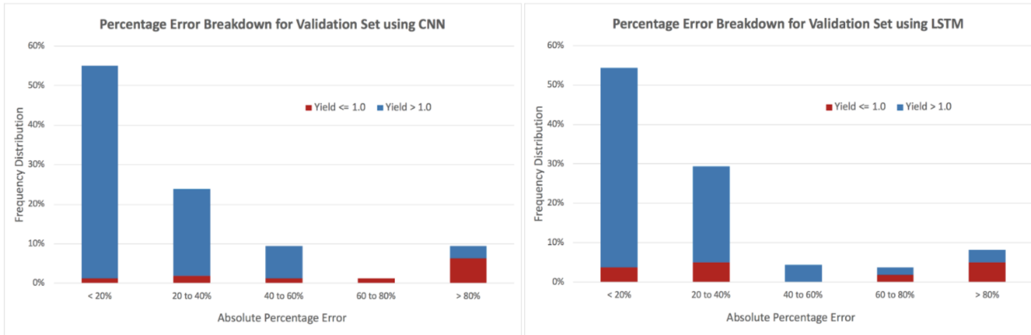

Figure 6: Best LSTM Model Results

6 Conclusion/Future Work

Crop yield forecasting using deep learning approaches can and should be extended to countries like India, where data availability and quality is especially challenging, and remote-sensing methods of estimation can be very cost-effective. Furthermore, shallow LSTM and CNN models performed fairly well on the limited data set, indicating the high accuracy that can be achieved if the government reports better yield data moving forward. Next steps for us include obtaining additional data for rice and other crops and to test similar models on them to assess how well the proposed techniques generalize. Lastly, a deeper error analysis to understand features of poorly classified images would be helpful.

7 Contributions

All team members worked cohesively on all areas of this project - literature review, data collection and cleaning, training models, analyzing results and preparing deliverables for milestones.

References

[1] “A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts.” International Journal of Forecasting, Elsevier, 16 Mar. 2016, `www.sciencedirect.com/science/article/pii/S0169207016000121`.

[2] “Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters.” International Journal of Applied Earth Observation and Geoinformation, Elsevier, 28 July 2005, `www.sciencedirect.com/science/article/pii/S0303243405000553`.

[3] Crop Yield Forecasting: Methodological and Institutional Aspects. Food and Agriculture Organization of the United Nations, Feb. 2016 `http://gsars.org/wp-content/uploads/2016/03/AMIS_CYF-Methodological-and-Institutional-Aspects_0303-web.pdf`.

[4] “Hunger and Food Security - United Nations Sustainable Development.” United Nations, United Nations, `www.un.org/sustainabledevelopment/hunger/`.

[5] Kondylis, Florence. “Measuring Yields from Space.” Impact Evaluations, 26 Oct. 2015, `blogs.worldbank.org/impactevaluations/measuring-yields-space`.

[6] Prasad, Anup K., et al. "Crop yield estimation model for Iowa using remote sensing and surface parameters." International Journal of Applied Earth Observation and Geoinformation 8.1 (2006): 26-33.

[7] Quarmby, N. A., et al. "The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction." International Journal of Remote Sensing 14.2 (1993): 199-210.

[8] Sabini, Mark, et al. “Understanding Satellite-Imagery-Based Crop Yield Predictions.” Stanford, Convolutional Neural Networks, 2017, `cs231n.stanford.edu/reports/2017/pdfs/555.pdf`.

[9] You, Jiaxuan, et al. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2017, www-cs.stanford.edu/~ermon/papers/cropyield_AAAI17.pdf.