Crop Yield Prediction using Deep Learning (SDG - 8)

Rushikesh Nalla 111971011 **Purvik Shah** 112045155 **Kaustubh Sivalenka** 111987216 **Sriram Gattu** 112007687

1 Introduction

Agriculture is one of the most important fields that has a major bearing on human sustenance and economic activity. There is immense scope for application of big data technologies to analyze trends and ameliorate processes in improving agricultural yield.

Through this project we tried to cover **Sustainable Development Goal - 8** that is **Decent Work and Economic Growth** as seen in Figure 2. The main problem being addressed in this project is: estimating crop yield (7) of soyabean as seen in Figure 1 based on satellite image understanding. Developing systems adept at crop yield estimation :forecast can help in various arms of the agricultural infrastructure by the way of optimizing inventory, control prices and preparing for food shortages. Agriculture sector employs 21.6 million people in USA making this sector one of the largest employers in the country.

Figure 1: Soyabean crop

Of late, there has been a deluge of satellite data (MODIS) (2) and crop yield statistics (NASS) that can be used to analyze crop yield trends over a spatio-temporal space (8). Big data technologies and machine learning pipelines constitute can be used to make apt analysis tools to combine satellite imagery and forecast crop yield trends (9). Said technological advances, when applied in real-world scenario translate into better productivity and loop in large number of skilled workforce contributing to economic growth and decent working conditions (better pay and secure working environment).

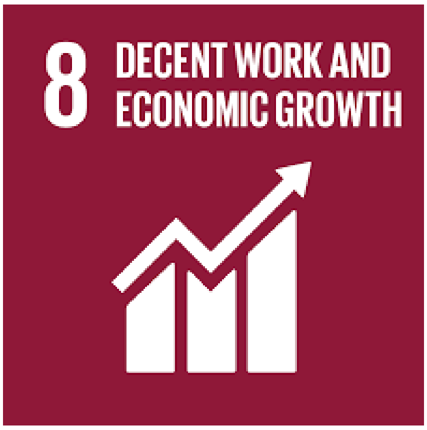

Figure 2: Sustainable Development Goal

2 Background

In the last decade, we have made huge strides when it comes to analyzing large amounts of satellite data. One novel application of the said data is that we can now predict the crop yield over years, months, specific counties and states. It is also one of the central challenges to Agricultural monitoring.

We have formulated crop yield prediction as a regression problem with satellite image features as independent variables and crop yield as the dependent variable in the context of ridge regression (6).

However, using satellite images, it would be a good idea to explore the trend of spatial feature evolution over time to forecast robust crop yield predictions. This idea was pursued using deep learning approach; LSTM (3) to

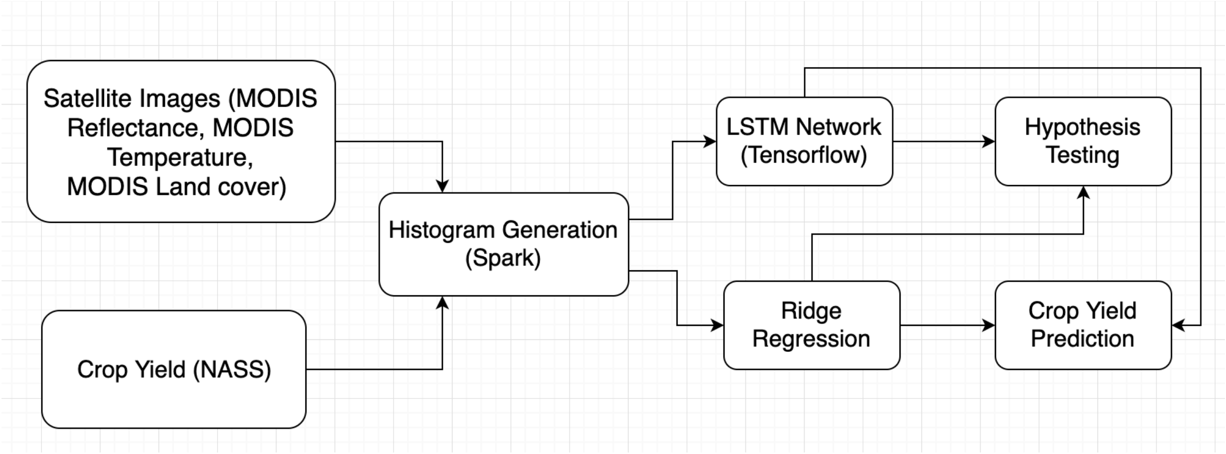
Satellite Images (MODIS Reflectance, MODIS Temperature, MODIS Land cover)
Crop Yield (NASS)
Histogram Generation (Spark)
LSTM Network (Tensorflow)
Ridge Regression
Hypothesis Testing
Crop Yield Prediction

Figure 3: Data Flow Diagram

predict future crop yield.

Our work can help those involved with Agricultural and Food industry economically. Big businesses can use this model to optimize price and inventory, government can prepare for food shortage, farmers can be informed of selling price if they know their regional yields and as a result shall attain a stable and secure agricultural operation.

3 Data

MODIS Satellite multi-spectral images are captured once in 8 days for each state and county in USA. In particular, three datasets from MODIS were taken (2) -

- Surface Reflectance which have 7 bands at 500m resolution
- Surface Temperature and Emissivity which have 2 bands at 500m resolution
- MODIS Land Cover at 500m resolution.

We collected data from 2003-2015 time period (inclusive). We have 9076 images in TIFF format (each 7.5MB). This was in total approximately 62GB in size. Dimension of each image is 48x48. The number of 8 day periods in a year are 46 and each image is captured using 9 bands. All images corresponding to a year were stacked together and the dimension was 48x48x414.

Corresponding crop yield for each state, county and year is also collected from NASS. We joined them and created a labelled dataset for prediction algorithm (described in next section).

4 Methods

Entire workflow of our pipeline is depicted in figure 3. We see that the Satellite multispectral image features are too large to train directly for a deep learning architecture. Hence, we first convert these images to histograms. Note that, while estimating crop yields by analyzing crop images, the exact location of pixel values hold less significance and can be permuted in any way and hence by this property of permutation invariance (4), conversion to histogram array would be a wise choice. Initially, we had to convert the format of satellite images from *.tif to *.npy so as to enhance the speed with which we pursue the content of these image files. Later, we fetched all the .npy files which were of the form (yyyy - countyid - stateid.npy). The data of each .npy file was a numpy array of size (48*48*414), where 48*48 signify the dimensions of the image, ”414” signifies the fact that for each year, an image was captured once every 8 days and each image encompassed 9-multispectral bands in total (i.e $414 = \lceil 365/8 \rceil * 9 = 46*9$).

For converting these images into histograms,

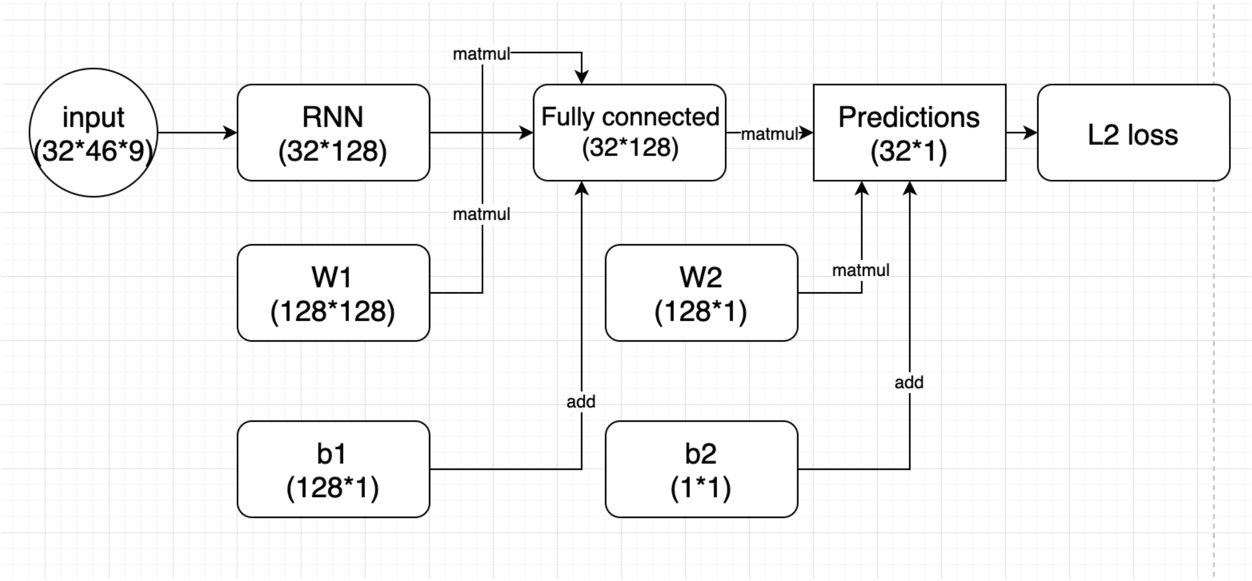

Figure 4: LSTM network overview

we distribute all the pixels based on their intensity into 32 bins. Then, we take a count of the number of pixels assigned to each bins and normalize them. The process of generating histogram is done in **Spark** so that we can do this work in parallel. The final spark RDD encompassed crop data as a list of elements of the form (year:[histogram array,crop yield]), which was eventually converted into a pandas pickle object of the very same format so that the the resultant pickle object encompassing crop yield data can be leveraged by both the LSTM and the ridge regression model.

After the histograms are generated, our pipeline splits in two ways. We predict Crop yield for each county using 1. Ridge Regression and 2. LSTM network coded in **Tensorflow**. After we've generated our results, we use hypothesis testing for validating our results. In our case, the hypothesis is, ”The features (Satellite images) can be reliably used for predicting the crop yield regions in USA.”

We fetch crop data from a pickle object of the form (year, histogram array of multispectral image, crop yield). Later, we built our train and test datasets in such a way that, if we had to predict crop yield for a particular year ”x”, then all the information gathered before that year was categorized as our training data and all the information relevant to the prediction year ”x” was categorized as our test data.

Ridge Regression: We leveraged scikit-learn module in our python script, so as to build a ridge model on our training data and hyper tune the regularization strength λ. Eventually after predicting the crop yields for various years, we reported the best regularization λ alongside the root mean square error incurred by the model whilst using the best λ. For practical purposes, L2 regularization was used for building an effective model.

Hyper parameter tuning for Ridge (best value): For ridge regression, we tuned L2 regularization strength λ of the model via a two phase approach.

Phase 1: Tuned λ with values from 0.1 to 1000000 by magnifying the value by a multiplication factor of 5, 10.

Phase 2: Observed a reasonable performance at λ = 500,so tried to explore various values in the vicinity of 500 (i.e., like 525, 550, 575, 625 etc) and **eventually it turned**

out that our model yielded it's best performance at λ = 575

LSTM network: As figure 4 suggests, our input is of size 32*46*9, where 32 is the number of bins, 46 in the number of 8 day periods in a year, and 9 is the number of channels. We create the LSTM Cell using Tensorflow's neural network library. After that, we apply the dropout wrapper on the cell. We stack multiple cells on top of each other using Tensorflow's MultiRNNCell.

After that, we use dynamic_rnn to feed the input data to the cell we've created. This results in a more compact graph. After that, we feed the output of the network to a fully connected layer which keeps the size of the output same. Then, we use another fully connected layer which gives our output same as our batch size. We store these as our predictions, and we calculate L2 loss from it after comparing them with our true values. We used Adam Optimizer for our network.

For training, we randomly pick a batch of data from our training set (as described above) and their corresponding crop yield. We run 1000 epochs for training, and we validate the data every 200 steps to check the root mean square error's progression. At the end of 1000 epochs we run the network to get the crop yield value for our test set (consisting of histograms for prediction year). We store the results and compare the results with true value of crop yields for different sets of analysis.

Hyper parameter tuning for LSTM and the best configuration: The hyper parameters we tuned when training LSTM were batch size, dropout probability, learning rate, and the size of the LSTM layer. We found out that the best results were achieved when was the **Batch size was 32, Drop out probability was 0.2, the Learning Rate 0.001, and the size of LSTM layer was 128.**

5 Evaluation / Results

Our primary intention was to demonstrate how high quality deep learning models like LSTM turn out to be a cut above the rest when compared to the conventional machine learning models like linear regression and ridge regression while analyzing huge corpus of satellite data whilst making predictions about crop yield.

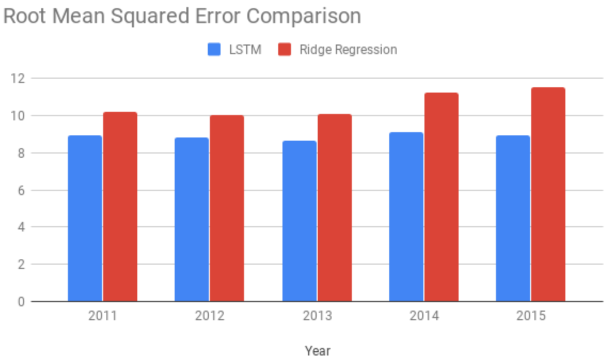

Figure 5: Comparison of root mean squared error resulted by both LSTM and ridge regression models

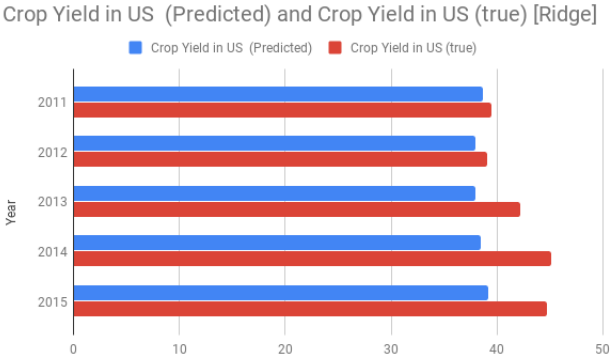

Figure 6: Comparison of true and predicted crop yields (Ridge Regression)

The bar chart in Figure 5 depicts the difference of the root mean square errors incurred by both LSTM and ridge regression models while predicting the crop yield for all the years ranging from 2011 to 2015 (inclusive). The fact that deep learning models perform better than linear models when fed with more training examples was bolstered by an apparent observation from the above figure in which as the years progressed, the root mean square error resulted by LSTM was much lower compared to the ridge regression. In other words, as the years progressed, the root mean square error gap between both the models magnified. The exact values of root mean square errors are provided in table 1.

The bar chart in Figure 6 gives a clear comparison of the mean value of the true crop yield and the mean value of the predicted crop yield by our ridge regression model for all the years ranging from 2011 to 2015 (inclusive). As the years progressed, the number of training examples scaled up and more spatio-temporal features came into picture and as a matter of fact, the performance of our ridge model degraded. Also, tuning L2 regularization parameter had no significant advantage after a certain point since L2 doesn't perform feature selection so as to capture interesting features.

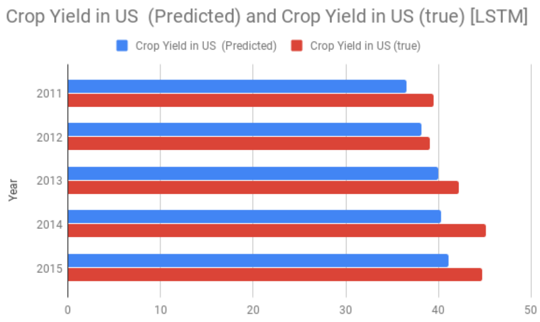

Figure 7: Comparison of true and predicted crop yields (LSTM)

The bar chart in Figure 7 demonstrates the comparison of the mean value of the true crop yield and the mean value of the predicted crop yield by our ridge model for all the years ranging from 2011 to 2015 (inclusive). It goes without saying that deep learning models are not good in dealing with less volumes of training data and a meticulous observation of the above two bar plots depicts the very same fact. That is, for the initial years (2011) where we had less training examples, the predictions made by our ridge model were slightly closer to the true predictions than LSTM. Though for the latter years (2012-2015), LSTM predictions were much closer to the true predictions since it was able to capture long term dependencies alongside the spatio-temporal features. The exact values of crop yield predicted by both LSTM network and Ridge regression are provided in table 2.

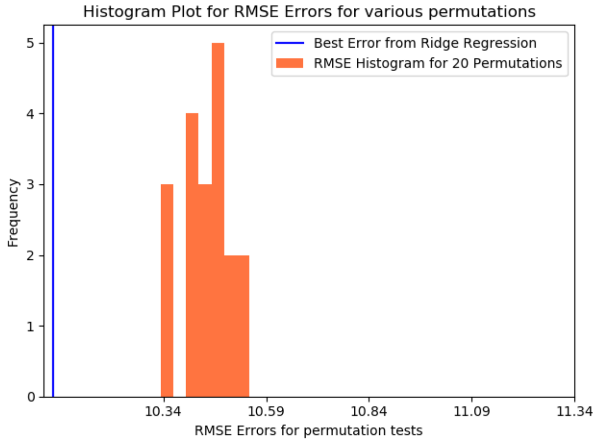

Figure 8: Permutation Test for ridge regression model

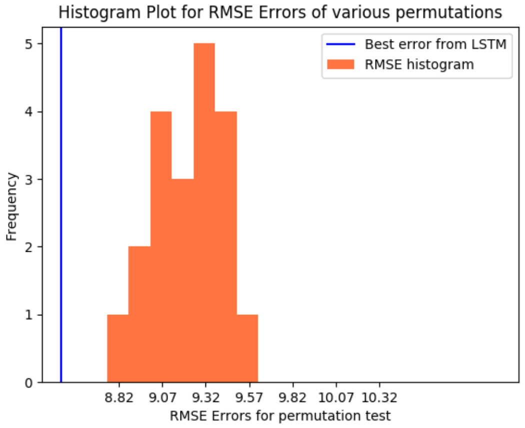

Figure 9: Permutation Test for LSTM model

In figure 8 and figure 9, we can see that the root mean square error obtained from the models trained on actual data is much lower as compared to permuted data i.e., crop yield values were shuffled so as disrupt the mapping of satellite images and corresponding crop yield. We obtained a p-value of 0 in both the cases.

Table 1: Root mean squared error for both methods

Year	LSTM	Ridge	Data [Training]	Data [Prediction]
2011	8.9212	10.18234	5248	601
2012	8.81759	10.05123	5849	631
2013	8.634521	10.07245	6480	603
2014	9.13451	11.22405	7083	621
2015	8.9456	11.5368	7704	558

Table 2: Crop Yield prediction and True Value for US

Year	True value	LSTM	Ridge
2011	39.4609	36.5286	38.632
2012	39.1082	38.1615	37.936
2013	42.2088	39.9307	37.977
2014	45.1212	40.3153	38.414
2015	44.7012	41.1035	39.162

6 Conclusions

- The LSTM approach to predicting crop yield provides a unique way to analyze satellite images for economic development of Agricultural and Food industry.

- We have done hypothesis testing, hyper parameter tuning on satellite data spanning multiple years, months, counties, and the country. We observe that the predicted crop yield is very close to the ground truth. This work can prepare those involved in the Agricultural/Food industries well and help them in their decision making.

- We also note that this network can be merged with Weather Sensors to obtain an even better real time prediction.

References

[1] Bolton, D. K., and Friedl, M. A., ”Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics.” In Agricultural and Forest Meteorology 173:74-84

[2] DAAC, N. L. 2015. ”The MODIS land products. http://lpdaac.usgs.gov.”

[3] Hochreiter, S., and Schmidhuber, J. 1997. ”Long short-term memory.” Neural computation 9(8):17351780.

[4] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, Stefano Ermon. ”Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data,” 31^{st} AAAI Conference on Artificial Intelligence (AAAI 2017)

[5] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). ”Google Earth Engine: Planetary-scale geospatial analysis for everyone. $Remote Sensing of Environment$”

[6] Arthur E. Hoerl Robert W. Kennard (1970). ”Ridge Regression: Biased Estimation for Nonorthogonal Problems”, Technometrics, 12:1, 55-67

[7] Anup K.Prasad, LimChai, Ramesh P.Singh, Menas Kafatos.”Crop yield estimation model for Iowa using remote sensing and surface parameters”. International Journal of Applied Earth Observation and Geoinformation.

[8] Ermon, S.; Xue, Y.; Toth, R.; Dilkina, B.; Bernstein, R.; Damoulas, T.; Clark, P.; DeGloria, S.; Mude, A.; Barrett, C.; and Gomes, C. 2015. Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In AAAI Conference on Artificial Intelligence.

[9] Jean, N.; Burke, M.; Xie, M.; Davis, M.; Lobell, D.; and Ermon, S. 2016. ”Combining satellite imagery and machine learning to predict poverty”. Science.