# Crop Yield Prediction using Deep Learning (SDG - 8)

{ RUSHIKESH.NALLA, PURVIK.SHAH, KAUSTUBH.SIVALENKA AND SRIRAM.GATTU}@STONYBROOK.EDU

8 DECENT WORK AND ECONOMIC GROWTH

## Introduction/Motivation

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 agriculutral yield.

There is a deluge of satellite data (MODIS) and crop yield statistics (NASS) that can be used to analyze crop yield trends over a spatio-temporal space.

Big data technologies and machine learning pipelines constitute apt analysis tools to forecast crop yield trends.

## 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 one of the central challenges to Agricultural monitoring.

Our work can help those involved with Agricultural and Food industry economically. In US, 21.6 million people are involved in the said industries. 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.

## Data

MODIS Satellite multi-spectral images are captured once in 8 days for each state and county in USA. Three datasets from MODIS were taken -
1) Surface Reflectance has 7 bands at 500m resolution
2) Surface Temperature and Emissivity has 2 bands at 500m resolution
3) MODIS Land Cover at 500m resolution.

We collected data from 2003-2015 time period. We have 9076 images (each 7.5MB) in total - approximately 62GB in size. Dimension of each image is 48x48x414, where 414 represents 46 (number of 8 day periods in a year) x 9 (bands).

Corresponding crop yield for each state, county and year is also collected. We joined them and created a labelled data-set for prediction algorithm.

## Conclusion

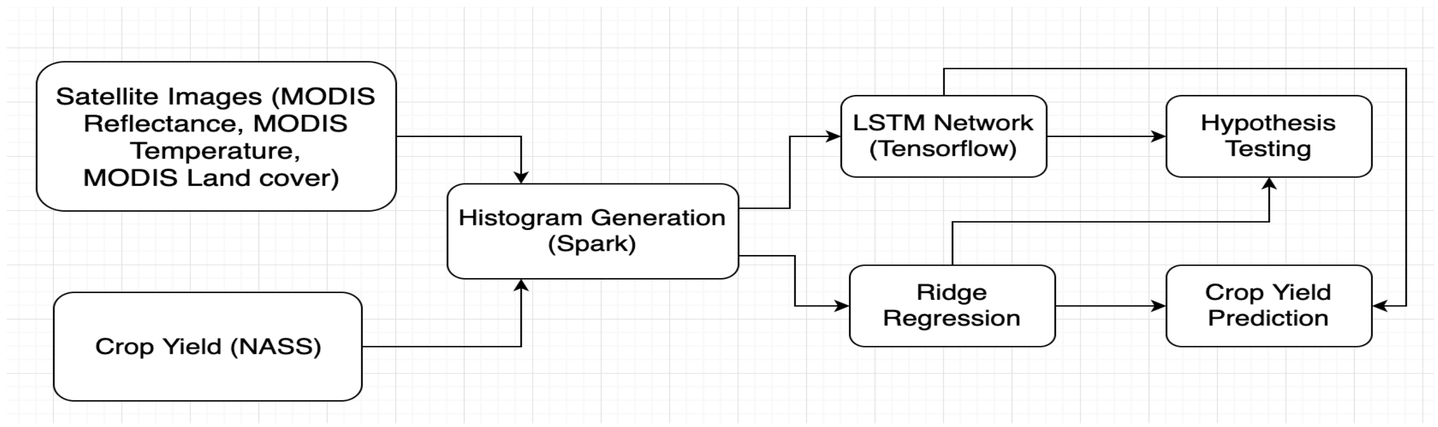
The LSTM approach to predicting crop yield provides a unique way to analyze satellite images for economic development of Agricultural and Food industry.

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.

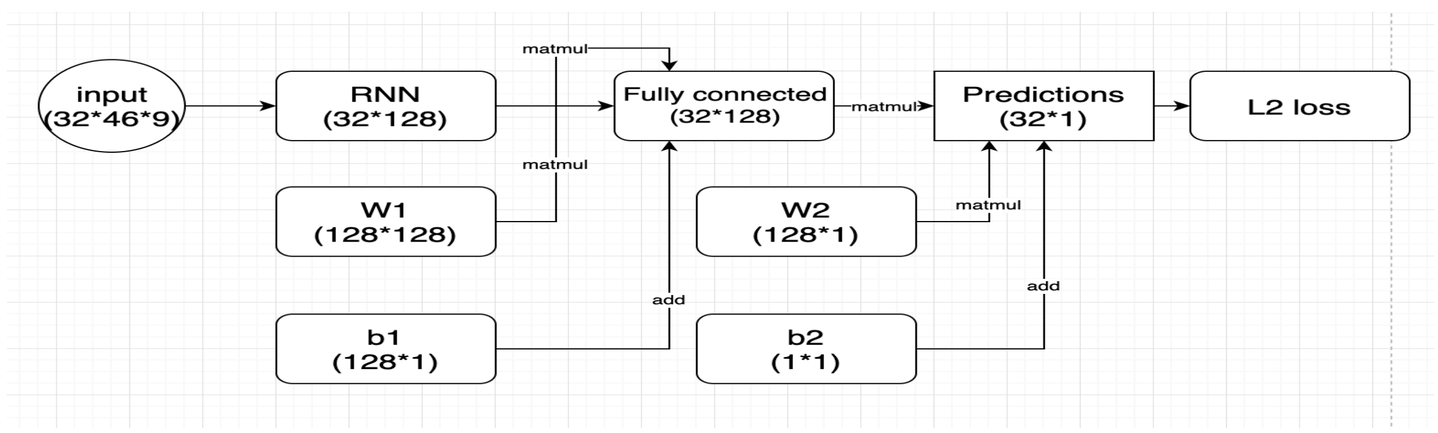
We also note that this network can be merged with Weather Sensors to obtain an even better real time prediction.

## Methods

**Data Flow Diagram**



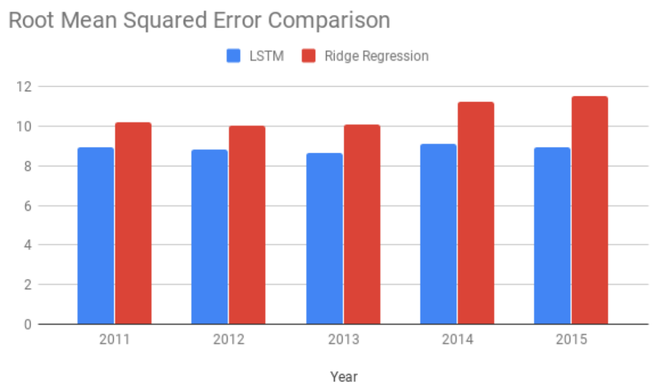

**LSTM Model in Tensorflow**



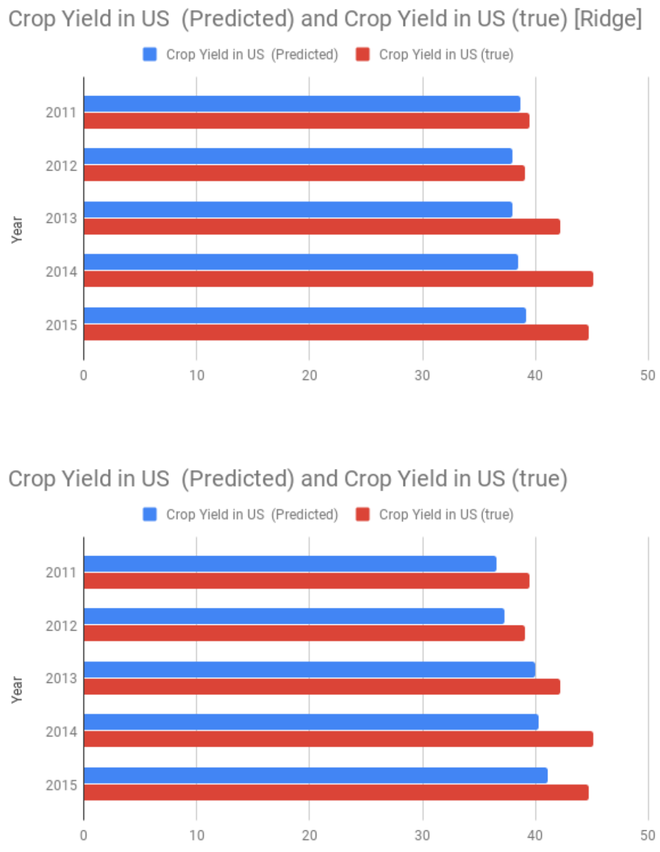

**Hyper parameters for LSTM (best value):** Batch size (32), Drop out probability (0.2), Learning Rate (0.001), LSTM Size (128), Weight Size(128)

**Hyper parameters for Ridge Regression (best value):** Regularization strength lambda $\lambda = 575$

## Evaluation/Results



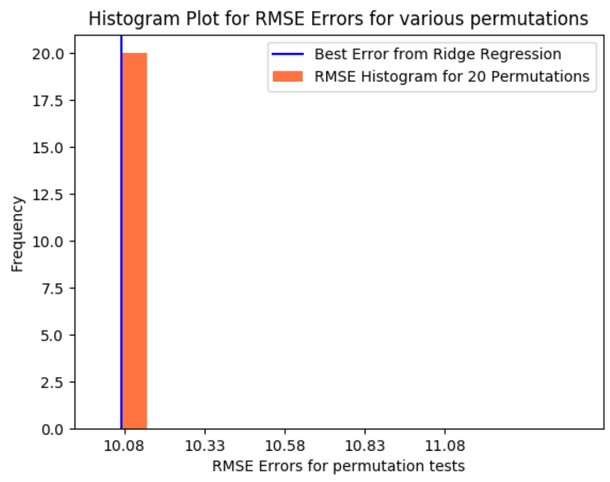

LSTM performs better than ridge regression as it captures spatial and temporal features.



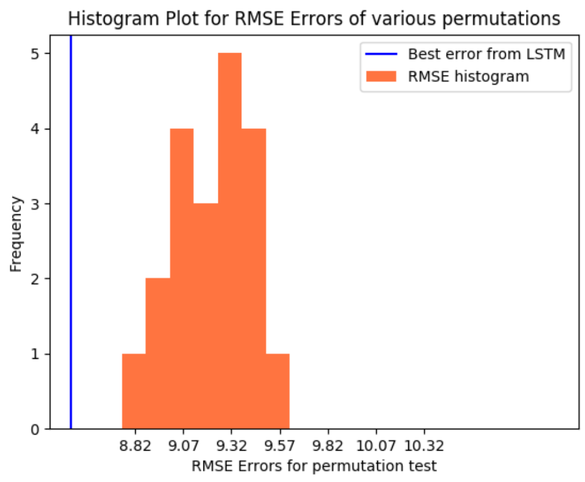

As the years progress, the difference between the actual and predicted crop yields kept decreasing for LSTM as compared to ridge regression. This is the fact that we had more data for latter years.

## Hypothesis Testing

**Permutation test for Ridge Regression**




**Permutation test for LSTM**




## 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):1735âĂŞ1780.

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