Pixel Count Based Yield Estimation Model, to Reduce Input feature required in Machine Learning System for Major Agricultural Crop

Atrayee Neog[1], Joydeep Kumar Devnath[1], Jishuraaj Nath[1], Raj Jyoti Sarma[1], Kalyan Lahkar[2]
[1]Assam Engineering College, Guwahati , INDIA
[2]Sumato Globaltech Pvt. Ltd. Guwahati, INDIA
atrayee.ent.aec@gmail.com

Abstract— ***Traditionally, the crop analysis and agricultural production predictions were done based on statistical models. However, with the climate of the world changing to drastic degrees, these statistical models have become very ambiguous. Hence, it becomes prudent that we resort to other less vague methods. Through a traditional model, user interacts primarily with a mathematical computations and its results and helps to solve well-defined and structured problems. Whereas, in a data driven model, user interacts primarily with the data and helps to solve mainly unstructured problems. At this point, enters the concept of Machine Learning. In this work we tried to find a new approach to reduce the input feature to reduce the processing power needed. We have attempted at predicting the agricultural outputs of rice production in an area by implementing a pixel count based classification machine learning model. Through this model, we tried to predict the approximate crop yield based on NDVI values analyzed for a particular season and area*.**

Keywords— Artificial Neural Network, Machine Learning, NDVI, Pixel count, Agricultural yield calculation, Low power computation..

1. Introduction

Agriculture, for decades, had been associated with the production of basic food crops, but in the present economic scenario agriculture plays a crucial role in the life of an economy. It is the backbone of our economic system. Agriculture not only provides food and raw material but also employment opportunities to a very large proportion of population. Traditional method of crop yield prediction models are numeric systems based on very strong statistical model. But due to drastic changing climatic conditions, systems are becoming very complicated and requires huge computational power. But with time as we are gathering more and more data, data driven modes become more meaningful and cost effective than numerical computational methods. To take the advantage of large data set available we are trying to co-relate two different types of data sets- high resolution airborne image and the yield estimation. We are trying to establish relational model using machine learning techniques.

Although yield estimation taking NDVI co-relation using machine learning has already been done and many commercial services are already available in the market, but the computation power required is quite high and can't be processed in mobile systems.

In this work we tried to find a new approach to reduce the input feature to reduce the processing power needed. We have attempted at predicting the agricultural outputs of rice production in an area by implementing a Machine Learning model. We tried to predict the approximate crop yield based on NDVI pixel values analyzed for a particular season and area.

2. Background Study

Study of satellite image has begun soon after the Landsat satellite was launched by NASA in 1972. Since the invention of Normalized Difference Vegetation Index (NDVI) in 1973, NDVI has been studied by many researchers. Dr. Robert Haas and et. al. had first found their ability to correlate, or quantify, the biophysical characteristics of the rangeland vegetation of this region from the satellite spectral signals[1].

NDVI is the ratio of VIS and NIR stand for the spectral reflectance measurements acquired in the visible (red) (VIS)and near-infrared regions (NIR) and it is calculated as

$$\text{NDVI}=\frac{NIR-VIS}{NIR+VIS} \quad (1)$$

Physically NDVI can be correlated as the color representation for particular vegetation. The reflected color from a leaf will represent the condition of the vegetation in the field. Figure1 shows reflectance of changing leaf appearance.

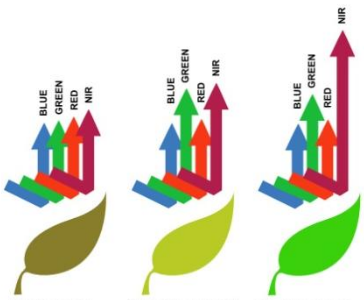

Figure. 1. Figure shows leaf's changing colour and corresponding reflectance from leaf. Source: Agribotix

The NDVI scale ranges from +1.0 to -1.0. Low NDVI value ie. 0.1 or less corresponds to areas with rock, sand, soil etc. Moderate NDVI value 0.16 to 0.5 corresponds to crops, grassland etc. and high NDVI 0.5 to 0.9 corresponds to very dense vegetation like forest or crop at peak growth stage [2].

Figure 2 shows NDVI scale with colour spectrum [1][7].

Fig. 2. Figure shows the NDVI scale with corresponding colour spectrum

Therefore, a multi band satellite image can be transferred to NDVI transformation and we will have a fairly good idea about approximate land use and land cover data. Figure 3 shows an example of a multiband image transformation to NDVI.

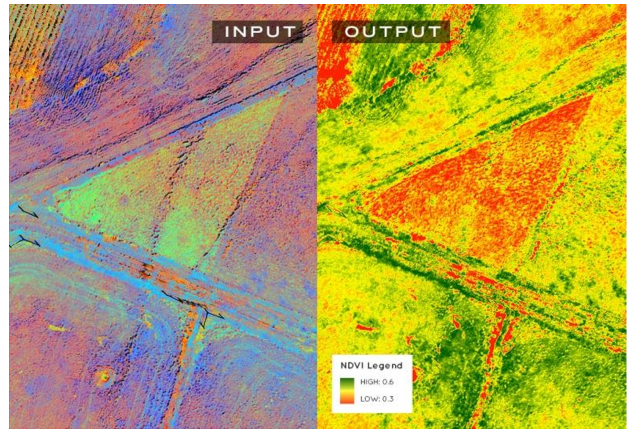

Figure. 3. Landsat 8 image of a paddy field to NDVI transf -ormation. Source:precisionhawk

M. S. Rasmussen had done operational yield forecast, where he established linear co-relation between yield collected and NDVI integrated during the reproductive period. Yield data could be available one month before the harvest [3].

N. A. Quarmby et. al. found encouraging results for operational crop monitoring. Yield for wheat, cotton, rice and maize crops has been estimated to a high degree of accuracy using a simple linear relationship between NDVI and yield. The estimates stabilize 50-100 days prior to harvest enabling an early assessment of crop yield to be made. However, input from an agro-meteorological model is recommended to modify the model during the grain-filling period of the wheat crop [4].

R. Singh and et. al has done crop yield estimation for all major crops in India. In an earlier study satellite spectral data has been used along with survey data to develop a more efficient post-stratified estimator of crop yield which suggested that with the use of satellite data along with crop yield data, it is possible to develop small area estimates of crop yield.

3. Methodology

In NDVI scale, NDVI value changes from -1 to +1 for a single NDVI image, number of input features become minimum 200 and additional 6 to 7 features from other supporting data like weather, season, crop type etc. Therefore, the model for the machine learning system becomes complicated and resource hungry. In this predictive system we have tried to implement machine learning technique to determine the yield form NDVI transformed image.

Therefore, to reduce the number of input features we have counted total number of Green, Yellow and Red (GYR) pixels in the image. This method considerably reduced the input features from two hundred to just three and hence the computation. To count the number of GYR pixels in the image we have made a Matlab program.

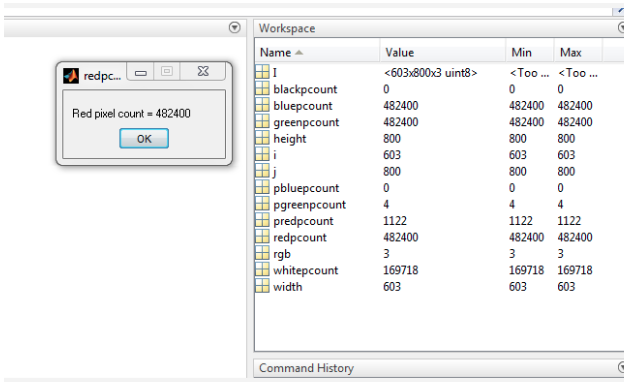

Figure. 4. Matlab program to count GYR pixels in the image

To reduce input feature data set we have converted the NDVI values to visual color pixels of Red, green, and yellow. Satellite images have been converted to the corresponding NDVI counterparts to generate an NDVI output with only three colors, Namely Green, Yellow and Red. The pixels belonging to these three colors are counted. Thus, the wide range of NDVI indexes have been mapped to only three features i.e. Red, Green and Yellow pixels. This marginally decreases the computational complexities as well as increases the computational speeds. Thus, it becomes very useful in case of real time data analysis or onboard calculations of data using smaller processors.

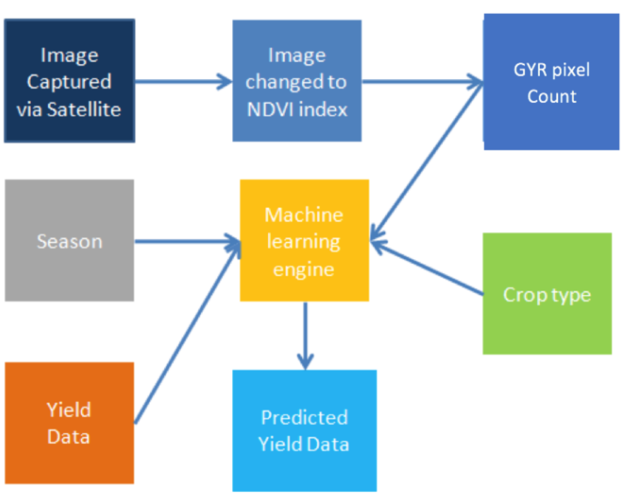

Figure. 5. The data flow process of the machine learning model

Datasets that have been used for this particular project have five features, namely: season, red, yellow and green pixel count and yield. The main purpose behind this project was to suggest a novel approach towards the prediction of crop yield for a particular region. The generated data has been processed with machine learning engine BigML. Regression model was used to train the machine learning engine.

Data set and analysis method

The data set was only for one crop. 80 percent of the generated data has been used to train the system and 20 percent has been used to validate the system. Season data kept variable, therefore we can adjust the season. GYR pixel count in total number and yield value for that particular crop is in quintal/Ha of land. The satellite image or drone based image has to in same scale, single crop and preferably of same location. Machine learning system was trained with 250 sets of data. Then it was validated with 50 data set. Table 1 shows features of the Machine learning model with their units.

I. **TABLE I**. TABLE OF FEATURES USED AND RESPECTIVE UNITS

Feature	Season	Red	Yellow	Green	Production
Unit	variable	Pixel count	Pixel count	Pixel count	Qt/Ha

To test the machine learning, season data and pixel count from the image has to be inserted. The BigML engine will return the probabilistic yield data for that crop.

4. Results and Discussions

This machine learning based model was found to be in accordance with the trained data and was found to be very promising for other crop data. The RMSE of the predicted data set has been found to be 232.68 Qt/ha. Therefore, we can expect a mean error of 232 quintal in every predicted value. Few of the test input data and their services are given shown in Table 2.

TABLE II. TEST DATA SET FOR DATA VALIDATION

Red	**Yellow**	**Green**	**Actual**	**Predicted**	**Error**
7657	745	782	7253	7456	-203
673	9874	662	9980	9835	145
782	9043	6788	14991	15474	-483
892	8932	6724	15243	15868	-625
672	8933	8945	652	756	-104

6732	5563	8752	11763	11568	195
563	223	892	411	489	-78
783	892	864	810	986	-176

5. Conclusion

The test result showed a very nice correlation using regression model between pixel count and estimated yield. In this study we have achieved a tolerance under error margin of 10%. This method has a future potential to process images in micro computer system in on-board computer system like drone

This method also opens future scope of implementing different machine learning model to find most optimized model so that can be run on onboard computer.

Further study is needed to see direct correlation of different colour pixels with yield.

References

[1] Myneni, Ranga B., et al. "The interpretation of spectral vegetation indexes." *IEEE Transactions on Geoscience and Remote Sensing* 33.2 (1995): 481-486.

[2] Lhermitte, Stefaan, et al. "Hierarchical image segmentation based on similarity of NDVI time series." *Remote Sensing of Environment* 112.2 (2008): 506-521.

[3] Rasmussen, Michael S. "Operational yield forecast using AVHRR NDVI data: reduction of environmental and inter-annual variability." *International Journal of Remote Sensing* 18.5 (1997): 1059-1077.

[4] 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.

[5] Singh, R. A. N. D. H. I. R., et al. "Small area estimation of crop yield using remote sensing satellite data." *International Journal of Remote Sensing* 23.1 (2002): 49-56.

[6] Labus, M. P., et al. "Wheat yield estimates using multi-temporal NDVI satellite imagery." *International Journal of Remote Sensing* 23.20 (2002): 4169-4180.

[7] Yuan, Fei, and Marvin E. Bauer. "Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery." *Remote Sensing of environment*106.3 (2007): 375-386.

[8] Müller, Andreas C., and Sarah Guido. "Introduction to machine learning with Python." (2017).