

## MACHINE LEARNING APPROACH FOR KHARIF RICE YIELD PREDICTION INTEGRATING MULTI-TEMPORAL VEGETATION INDICES AND WEATHER AND NON-WEATHER VARIABLES

Aditi Chandra\*<sup>1</sup>, Pabitra Mitra<sup>1</sup>, S. K. Dubey<sup>2</sup> and S. S. Ray<sup>2</sup>

<sup>1</sup>Dept. of Computer Science & Engineering, IIT, Kharagpur, India – (aditi.chandra.iitkgp, pabitra)@gmail.com

<sup>2</sup>Mahalanobis National Crop Forecast Centre, DAC&FW, New Delhi 110 012, India - sunil.dubey86@gov.in,  
shibendu.ncfc@nic.in

Commission III, WG III/10

**KEY WORDS:** Artificial Neural Network, Random Forest, NDVI, Kharif Rice, Yield Prediction, Purulia, Bankura,

### ABSTRACT

The development of kharif rice yield prediction models was attempted through Machine Learning approaches such as Artificial Neural Network and Random Forest for the 42 blocks covering 13,141 sq km upland rainfed area of Purulia and Bankura district, West Bengal. Models were developed integrating monthly NDVI with weather and non-weather variables at block-level for the period 2006 to 2015. The model correlation obtained was 0.702 with MSE 0.01. Though the weather variables vs NDVI models are quite satisfactory, NDVI vs kharif rice yield models however, show relatively less correlation, about 0.6 revealing the requirement of varied additional farmer-controlled inputs. Development of NDVI vs crop yield models for different crop growth stages or fortnightly over a larger data set with selective adding of weather and non-weather variables to NDVI would be the most appropriate.

### 1.0 INTRODUCTION

In India, rice is one of the most important crops and contributes significantly to national economy. It is grown in about 4.34 lakh sq km area spread in almost all the States and constitutes about 41% (2015-16) of the total food grains output. The rice yield has increased over the years owing most to the agriculture technology initiatives and farmers' skill. However, the long-term climate change and the complexities of seasonal weather variability have induced an inevitable uncertainty in yield. It is more so for kharif rice yield in the rainfed area which is most vulnerable to southwest monsoon rainfall- the main source of water. In India, such rainfed area occupies about 75 million hectares, i.e., about half the net sown area of 139.9 million hectares. In this regard crop-yield prediction before harvest is required by planners as well as farmers. Within the growing season, as early as possible, farmer wants to know reliably how much yield he can expect so that to enhance yield, if needed and scope exists, necessary additional inputs can be made. The yield and advance estimates of crop production are computed through a large number of crop cutting experiments. An improvement in crop-area estimates has been made with much better spatial representation under FASAL (Forecasting Agriculture using Space, Agrometeorology and Land based observations) scheme wherein an analysis of multi-temporal remote sensing data through satellite images is incorporated with the statistical and crop simulation models. The satellite data-based crop prediction is being done in India by Mahalanobis National Crop Forecasting Centre (MNCFC), New Delhi.

Satellite image data can provide vital information on crop inventory, classification, biomass accumulation during different phenological stages, growth rate, spatial yield variability, near-real time monitoring and the best pre-harvest stage information for useful yield prediction (Groten, 1991, Mkhabela et al), and ultimately the yield. The red and near infra-red regions reflectance based Normalized Difference Vegetation Indices (NDVI) derived from multi-temporal satellite images provide a good representation of plant photosynthetic capacity and efficiency and can be related to crop yield. As pointed out by Kasampalis et al (2018) the high temporal frequency and broad spatial coverage of the satellite data at low cost has made it a preferred choice. The NDVI based crop yield prediction is done through regression model and could be integrated with statistical and crop simulation models. A wealth of information on yield predictions using historical NDVI data for different crops from several countries is available. Some of the studies using NDVI for yield predictions are by Dubey et al (2018) for sugarcane, Lopresti et al (2015) and Pantazi et al (2016) for wheat and Chang et al (2005) and Huang et al (2013) for rice yield. Rice yield prediction at regional scale in Indo-Gangetic river basin through NDVI is attempted by Cai & Sharma (2010). Nevertheless, the application of satellite data in rice yield prediction is limited compared to other crops (Huang et al, 2013). For improving the performance of crop prediction model Prasad et al (2006) add soil moisture and weather parameters influencing the yield to NDVI while Liang et al (2004) integrate dynamic crop simulation model with NDVI.

This paper focus on developing kharif rice yield prediction models for rainfed upland area through machine learning techniques integrating block (sub-district) level NDVI with weather and non-weather variables.

---

\* Corresponding author

In agriculture sector quite a large volume of data at varied temporal and spatial scales is generated every year on a good number of weather and non-weather variables on which yield depends and the satellite based remote sensing data. It provides ample scope for the applications of data driven advanced techniques of machine learning to analyze and may prove to be an efficient approach in crop yield prediction under the climate change scenario. The machine learning is a data driven self adaptive automated method, not requiring any knowledge of the physical relations or mechanisms that produces the data. The machine is provided with an access to data to learn and validate by itself. It learns the functional inter-relations from training and validates for the predictions. Because of this, machine learning techniques are quite robust and easy in analyzing the patterns and rules within large data sets of multiple predictor-variables having nonlinear relations with the target variable. Out of several machine learning techniques, the artificial neural networks (ANN) has been used considerably in crop-yield predictions. The use of remotely sensed data and deep learning in estimating the crop yield are in practice (Kuwata and Shibasaki, 2015 and 2016). The remote sensing data and Gaussian Process component are used to model the spatio-temporal structure of the data in predicting crop yields (You et al, 2017). The multivariate regression and neural network techniques have been combined on multi-temporal satellite data to estimate the crop yield (Li et al, 2007). Nitze et al (2012) compare the applications of Random Forest (RF), ANN and support vector machine (SVM) in crop type classification utilising remote sensing data.

## 2.0 MATERIALS AND METHODS

In general, the prediction models are developed at district, state or country scale. The present study attempts for kharif rice yield prediction at block (sub-district) scale in the rainfed upland areas of West Bengal, India; the reason being the variations in weather and yield amongst the blocks. In the study the weather and non-weather predictor variables are integrated with NDVI data at block level to predict the yield. The NDVI data being available from 2006 onward, and yield data available only up to 2015, the prediction models are generated based on 2006-2015 data.

### 2.1 Study Area

The study area is rainfed uplands of Purulia (22.7° to 23.7° N, 85.8° to 86.9° E) and Bankura (22.6° to 23.65°N, 86.9° to 87.75°E) districts, located in the westernmost part of West Bengal (Fig.1). It occupies a total geographical area about 13,141 sq km and has 42 administrative block (sub-district) development units; Purulia: 6259 sq km, 20 blocks and Bankura: 6882 sq km, 22 blocks. As per rainfed area prioritization, Purulia and Bankura districts are respectively at 1<sup>st</sup> and 9<sup>th</sup> rank in West Bengal and 112<sup>th</sup> and 292<sup>nd</sup> rank among the rainfed districts of India (NRAA, 2012). Almost the entire area of Purulia district and the western half of adjacent Bankura district (15 blocks) are in hard rock and marginal alluvial terrain. It is dominated by granite gneiss and schist rock and forms the north-eastern part of Chhotanagpur plateau. It is mainly covered by residual soil ranging from gravelly loam to fine loam derived from parent

rock weathering and also with laterite capping in the upland areas. The eastern part of Bankura district is occupied by sedimentary rock formations and covered by alluvial soil. The soil texture type varies from block to block and within the block also. In general, the soils in Purulia district and western part of Bankura district are of low fertility. The general variation in topographic elevation is from 80 to 312 m above mean sea level, giving rise to a broad local classification of land as southwestern and western upland in Purulia district and eastern and southeastern low land in Bankura district. There are several isolated hills and dissected plateau and residual hills presenting an undulating topography. The district is drained mainly by 3 rivers, viz., Kangsabati (Kasai), Damodar and Dwarakeshwar. About 617 sq km area in the western and north-western parts of Purulia district and 1464 sq km area distributed evenly in entire Bankura district is occupied by forest. The area gets southwest monsoon rains in the months of June to September. During 2006-2014 the monsoon total rainfall varied from 394.3 mm (2010) to 1681.8 mm (2007) in 20 blocks of Purulia district and from 614.9 mm (2010) to 1628.3 mm (2007) in 22 blocks of Bankura district. The variations in rainfall, temperature, solar radiation, relative humidity and NDVI values are shown in Table 1 and kharif rice yield in Table 2.

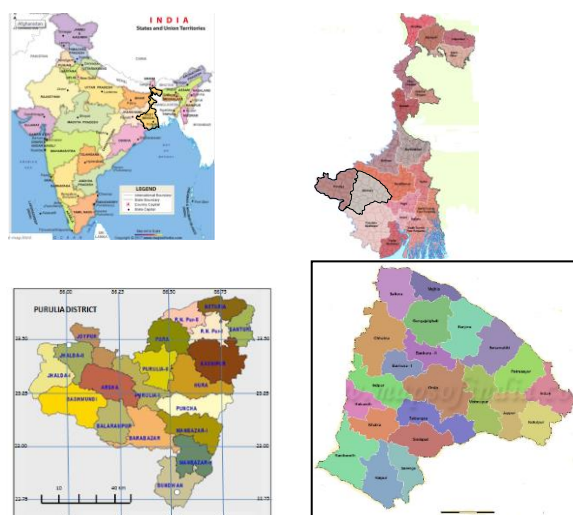


Figure 1. Map of the study area: Purulia and Bankura districts, West Bengal with block boundary and grids.

### 2.2 Weather and Non-Weather Variables Affecting Yield

In Purulia and Bankura districts, 1000 to 1500 mm water requirement of kharif rice grown during the monsoon season is met by rains. The deficient rainfall during summer of 2010 was critical. The monsoon rainfall was 500-600 mm less than the normal that resulted in a decline in kharif rice yield by 25-35% and crop area by 40%. Besides, kharif rice yield is vulnerable to onset and withdrawal of monsoon, temporal distribution of rainfall in the growth stages and number and duration of dry spells. Erratic distribution of rainfall affects the crop yield. Heavy rainfall during transplantation and deficient rainfall during growth may lead to yield loss. Since Purulia and Bankura districts generally get the required rainfall, the variation in rainfall is not always a constraint for

low to moderate kharif rice yield of 1500-2500 kg per hectare which is generally maintained through farmers' input and like fertilizer, seed variety, pesticides etc and technologies. The variables like temperature, solar radiation, relative humidity also play important role in obtaining the desired yield. The impact of decreasing rainfall is not much evident over the years, as it is masked by a general increasing trend in kharif rice yield due to switching over to modern variety rice during 90's and continued changes in agricultural technology over the years.

The average temperature required for kharif rice throughout its growing stages ranges from 20° to 33° C. Maximum temperature (TX- day temperature) in the range of 20° to 33° C enhances the growth (Welch et al, 2010). In Purulia and Bankura districts TX ranges from 32.8 to 27.3 °C in the months of August-October. An increase in TX within the range has a positive impact during vegetative and ripening stages. Yield of rice is affected by minimum temperature (TM-night temperature) also. In these districts TM ranges from 17.1 to 25.9° C. Peng et al (2004) show that the impact of variation in TM is more pronounced compared to that of TX. Rice yield is positively related to biomass production which is the product of intercepted solar radiation by the canopy and the radiation use efficiency (Huang et al, 2016). Yield of rice reduces with reduced solar radiation, mainly in the latter part of the season. Large solar radiation has positive impact during ripening and negative impact during vegetative phase and it brings in precision in the prediction. Rice requires high relative humidity of 80-85%. Low relative humidity of around 60% or less decreases the moisture content and reduces the yield. However, the impact of variations in relative humidity at different phases of growth is different. Higher relative humidity has a negative impact during ripening phase. Rathnayake et al (2016) reveal that higher relative humidity and higher temperature during flowering stage has a negative impact.

Rice can be grown on a variety of soil in rainfed lowland and upland. The silty, clayey, clay-loamy and loamy soils are suitable; while red and lateritic soil, having relatively less water retention capacity, are generally not suitable. In the upland hard rock terrain of Purulia and Bankura districts, the soils are Red and lateritic in major part (Adhikari et al, 2011). These soils have coarse texture and low water holding capacity and therefore the fertility or productivity is low in general (Bera, 2018). A variety of rice seeds is in practice in West Bengal. The conventional, short and long duration, hybrid and high yielding variety seeds are used depending on the climatic conditions and the soil type. In Purulia district the HYV area has increased from 87.35% in 2004 to 96.58% in 2014 (Bureau of Applied Economics & Statistics and Economic Review, Govt. of West Bengal). Fertilizers are required to balance the soil nutrients or remove the deficiencies and therefore judicious and timely use of fertilizer in required quantity and rate has a major role in increasing the rice yield. In Purulia and Bankura districts the kharif season NPK fertilizer consumption during 2006 to 2014 varied from 28.26 to 61.16 kg/ha and 85.47 to 246.41 kg/ha respectively. The maximum consumption rate of 61.16 kg/ha in Purulia district and 246.41 kg/ha in Bankura district was in 2010 which was a deficit rainfall year. Also the

general consumption rate of NPK fertilizer in kharif season in Bankura district was almost three times that in Purulia district and kharif rice yield in Bankura district was 1.12 to 1.44 times that in Purulia district. The value of NDVI quantifies the vegetation. It is computed from the difference between the near infrared (NIR) and red light (Red) band reflectance. NDVI value ranges between -1 and +1. Negative values are associated with water bodies and positive values are associated with vegetation. NDVI values close to +1 are indicative of dense green healthy vegetation. Multi temporal NDVI data is obtained from MODIS (Moderate Resolution Imaging Spectro-radiometer) Terra Satellite data. In Purulia district NDVI ranges from 0.223 (July) to 0.778 (Sept) and in Bankura district it ranges from 0.236 (June) to 0.813 (Sept). A reduction in NDVI observed in the month of Sept/Oct in some of the years could be due to cloud cover. Compared to Purulia district in Bankura district an increase in NDVI (Sept) by 0.35, fertilizer consumption by 3 times and yield by 1.12 to 1.44 is observed.

### 2.3 Data -Source, Processing & Visualisation

To build the models we collected historical daily weather data, viz., rainfall (RF), minimum and maximum temperature (TM, TX), solar radiation (SR) and relative humidity (RH), for the period 1983-2015 as per the availability, through 2 sources. From India meteorological Department (IMD) we got RF data at 0.25° x 0.25° grid. The 6259 sq km geographical area of Purulia district having 20 blocks, approximately within 22.75° N to 23.75° N latitudes and 85.75° E to 87.0° E latitudes can be represented by 14 such grid points (Fig.1). Similarly, Bankura district having 22 blocks, approximately within 22.5° N to 23.75° N latitudes and 86.75° E to 87.75° E can be represented by 14 grid points. The daily TM, TX, SR and RH data at the grid points were obtained through NASA website. The daily data were preprocessed to obtain monthly total RF and other data averaged at the 14 grid points. The fertilizer consumption data for kharif season for each year for the whole district was obtained from Fertilizer Association of India, Kolkata. The block level fertilizer consumption for each year was proportionately approximated from district level consumption through block level kharif rice cultivated area. The percent area with High Yielding Variety (HYV) seed at block level was approximated from district and state level yearly data. The data on soil type could be generated from the soil map of the district. The dry crop (clean rice) yield for the district and blocks was obtained from agri-statistics, Govt. of India website and Department of Planning, Statistics and Programme Implementation, Government of West Bengal. The block level monthly NDVI data were obtained from Mahalanobis Nation Crop Forecast Centre (MNCFC), New Delhi.

### 2.4 Machine Learning Approaches

The Machine Learning techniques, viz., Artificial Neural Network (ANN) and Random Forest Regression (RFR) were used to develop models at block level of Purulia and Bankura districts using the block level rainfall, minimum maximum temperatures, solar radiation and relative humidity and NDVI

data for the months of June to October over the period 2006 to 2015.

## 2.5 Methodology

The study started with 20 blocks of Purulia district, followed by 22 blocks of Bankura district and then combination of these districts. Fertilizer, soil type and seed variety (HYV) were included as additional predictor variables at the initial stage, but later on soil type and HYV were taken out as these were not found having any impact on the model accuracy. The study started with yield as target variable and different combinations of predictor variables, viz., weather variables, weather variables and fertilizer, NDVI, weather variables and NDVI and finally weather variables, fertilizer and NDVI. The development of models started with ANN without feature selection followed by ANN with feature selection through Boruta algorithm (Kursa et al, 2010; Rudnicki et al, 2015) and with variables of importance obtained through Random Forest (Breiman, 2001). Feature selection becomes a necessity when the number of predictor variables (features) is large. In the present case dealing with 5 monthly data for 5 weather variables, ANN models gave relatively poor correlation. It required identification of all-relevant variables, i.e., to reduce the number of variables or reduce dimensionality. The Boruta box plot is shown in figure 2. A comparison of model accuracy was made to select the model for prediction (Table 3). The open source R language was used.

The most commonly used neural network architecture, the feed forward back propagation neural network was used, which requires less memory and is relatively fast. It is a supervised learning method using known outputs to train the model. In back propagation learning process the error in output propagates backward and the weights are adjusted to minimize the error. Basically, ANN models attempt to search for relationship between the input and output variables to define a formula which can be used for predictions. To improve the prediction, back propagation is used to modify the weights. The sequence of operation includes preparation of appropriate and adequate data sets, training of the neural network and testing. For training randomly selected 75% of the data sets were used and the remaining 25% data sets were used for testing the accuracy of the model generated. The correlation and mean square error (MSE) were computed to evaluate and ascertain the model performance. In R Neuralnet package the training of the neural network model was done using resilient back propagation algorithm. Only one hidden layer was used. ANN models were trained for a varied number of hidden layer neurons. The best result was obtained with seven hidden neuron when only weather variables were used and five when weather and non-weather variables were used. The final number of hidden nodes was decided after making several runs with different numbers of hidden nodes. On an average 50 model runs were made to compute the average correlation and Mean Squared Error (MSE).

## 3.0 RESULTS AND DISCUSSIONS

The models developed are summarised in Table 3. and discussed below. The correlations are averaged over 50

model runs. The model MSE, in general ranges from 01 to 0.03.

### 3.1 Weather Variables vs Yield Models

Yield prediction models were developed for the blocks of Purulia and Bankura districts for the period 1992 to 2015 (24 yrs) using 5 monthly weather variables only. The Purulia model correlation is 0.669. With Boruta feature selection the model correlation improved to 0.723 with MSE 0.01. Adding fertilizer does not improve the model accuracy. Since the NDVI data are available from 2006 onwards, the weather variable-yield models were also run with 2006-2015 data. With this the ANN model is non-convergent. The correlation of Purulia ANN model with feature selection is 0.674 to 0.698. In case of Bankura district the ANN models with feature selection show a correlation of 0.515. Subsequently, Purulia-Bankura combined ANN models developed. The combined model correlation is 0.549 and with Boruta and random forest the correlation improved up to 0.649. Adding fertilizer improves the correlation further up to 0.682. Unexpected but significant finding is that in Boruta feature selection or random forest variable importance the rainfall has a poor ranking. The reduction in correlation for 2006-2015 models compared to 1992-2015 is due to less number of observations.

### 3.2 Weather Variables vs NDVI Models

The models developed with monsoon monthly weather data as predictor variables and NDVI as the target variable for the period 2006-2015 for Purulia and Bankura districts show a correlation of 0.836 with MSE 0.0135 indicating high dependency of NDVI on weather variables. The Purulia-Bankura combined model show a correlation of 0.821 with MSE 0.017. The predictor variable importance ranking obtained from random forest regression in decreasing order is TM, SR, RH, TX and RF. Like weather variable vs yield models, here also, the RF has relatively poor variable importance.

### 3.3 NDVI vs Yield Models

The ANN models for 5 monthly (June to October) NDVI values as predictor variables and crop yields as target variable yielded the following results. The models based on NDVI only show a correlation of 0.431 to 0.571 for Purulia district and 0.436 to 0.455 for Bankura district. In case of Purulia district marginal improvement in correlation is observed when July to October NDVI is considered while for Bankura district the correlation improves marginally with August to October NDVI. Combining Purulia with Bankura improves the correlation up to 0.604 for July to October NDVI. The improvement in correlation for combined model could be attributed to larger observations (1000) and secondly, there are several farmer-controlled variables which are not taken into account. In a related block level study based on growth stage wise data for Purulia district, the RFF models on weather variables vs crop yield (without NDVI) showed a correlation of 0.766 with predictor weather variable importance ranking in decreasing order was Solar Radiation (Stage III : August 13 to September 15), Relative Humidity

(Stage III) and Maximum Temperature (Stage IV : September 16 to October 31). While in ANN Models the adding of October NDVI to input variables as Solar Radiation and Relative Humidity of growth stage III showed a minor improvement in correlation (0.80), the adding of September NDVI instead of October NDVI lessened the correlation

### 3.4 Weather Variables & NDVI vs Yield

The models with weather and NDVI as predictor variables have better correlation for Purulia district compared to NDVI vs yield models. ANN models with random forest 7 predictor variables of importance show a correlation of 0.685 and the models on 10 relevant features selected through Boruta algorithm have a correlation of 0.633. For Bankura district, the correlation values are comparatively less of the order of 0.432 to 0.542. The ranking of features is not consistent for these districts. For Purulia-Bankura district combined ANN models also, the correlation with Boruta feature selection and random forest variable importance ranges from 0.651 to 0.663. A comparison of performances of models on NDVI-Yield with the models on weather variables, NDVI –Yield reveals that adding weather variables to NDVI improves the correlation though marginally. Secondly, the relation between weather variables to yield and NDVI to yield varies spatially from Purulia to Bankura district. A point to be mentioned here is that while the former is totally in hard rock terrain, the latter is in hard rock as well as marginal alluvial terrain.

### 3.5 Weather Variables, Fertilizer and NDVI vs Yield

Adding of fertilizer and weather variables as a predictor to NDVI for the individual districts does not improve the correlation. However, the Purulia-Bankura combined models show an improvement in correlation up to 0.702 with MSE 0.010. The combination Fertilizer, Relative Humidity (August-September) and NDVI (September-October) Maximum Temperature appears to be the most appropriate to explain the yield variability. As observed the variable importance varies from Purulia to Bankura and with larger data set of Purulia and Bankura combined.

## 4.0 CONCLUSIONS

The study reveals that for rainfed kharif rice yield prediction NDVI alone may not yield the desired model. Adding of weather and non-weather predictor variables to NDVI values improves the model accuracy up to 0.7. Instead of monthly data the models based on different time granularity such as crop growth stage wise or fortnightly NDVI and weather variables are likely to have much better accuracy. The seed variety and soil texture type has shown little relevance. The present study is based on 9 years' data only. A large data size would bring in better accuracy and stability. Block level study is appropriate. However, the grouping or selection of blocks has to be done after data visualization and anomaly verification of NDVI and weather variables and non-weather variables even at district level. In ANN the time-sequence of variables has not been considered. The combination of Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) which has the potential to model the time sequential effects

of weather variables and NDVI would yield the desired prediction model for kharif rice yield.

## ACKNOWLEDGEMENTS

Authors are thankful to Professor Abhijit Mukherjee, IIT, Kharagpur, the Director, Department of Planning, Statistics and Programme Implementation, Government of West Bengal, to the Head, Fertilizer Association of India, Regional Office, Kolkata and to Shri Amit Rabha, Analyst, MNCFC for providing the relevant data and useful discussions. The first author is also grateful to Mrs Priyanka Singh and Shri Arnav Chandra for fruitful discussions on crop yield held time to time.

## REFERENCES

- Adhikari, B, M. K. Bag, M. K. Bhowmick & C. Kundu (2011), *Status Paper on Rice in West Bengal*, Rice Knowledge Management Portal (RKMP), Directorate of Rice Research, Rajendranagar, Hyderabad 500030, <http://www.rkmp.co.in>
- Bera, S (2018), Delineation of Ground Water Prospect Zone in hard rock terrine part of Purulia District of West Bengal-India using Remote Sensing & GIS Approach, IJRSCEIT, Vol. 3, No. 1, pp 1361-1370
- Breiman, L. (2001) *Random Forests, Machine Learning*, Kluwer Academic Publishers, The Netherlands, Vol. 45, pp 5-32
- Bureau of Applied Economics & Statistics, District wise Estimates of Yield Rate and Production of Nineteen Major Crops of West Bengal (1995-2006, 2010-2015), Department of Statistics and Programme Implementation, Government of West Bengal
- Cai, X. L. And B. R. Sharma (2010) Integrating remote sensing, census and weather data for an assessment of rice yield, water consumption and water productivity in the Indo-Gangetic river basin, *Agricultural Water Management*, Vol. 97, pp 309-316
- Chang, K.-W., Y. Shen and J.-C. Lo (2005) Predicting rice yield using canopy reflectance measured at booting stage, *Agron. Journal*, Vol. 97, pp 872-878
- Dubey, S.K., A. S. Gavli, S.K. Yadav, S. Sehgal and S.S. Ray (2018) Remote sensing –based yield forecasting for sugarcane (*Saccharum officinarum* L.) crop in India, *Jour. Indian Soc. Remote Sensing.*, 46(11):1823-1833.
- Groten, S.M.E. (1991) NDVI-Crop monitoring and early yield assessment of Burkina Faso, Paper presented at the UN/FAO/UNEP Seminar on Remote Sensing for Decision Makers, Nairobi, Kenya, 26-31 August 1991, *Int. Jour. Remote Sensing*, Vol. 14, pp 1495-1515
- Huang M, Shan, S, Cao F &, Zou Y (2016), The solar radiation-related determinants of rice yield variation across a

wide range of regions, *NJAS-Wageningen Journal of Life Sciences*, Vol. 78, pp 123-128

Huang, J., X. Wang, X. Li, H. Tian and Z. Pan (2013) Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR, *PLoS ONE* 8(8): e70816.  
<https://doi.org/10.1371/journal.pone.0070816>

Kasampalis D. A., T.K. Alexandridis, C. Deva , A. Challinor , D.Moshou and G. Zalidis (2018) Contribution of Remote Sensing on Crop Models: A Review, *Jour. Imaging*, Vol.4 (52), doi:10.3390/jimaging4040052.

Kursa, M B., A Jankowski, W R. Rudnicki, 2010, Boruta – A System for Feature Selection, *Fundamenta Informaticae*, Vol. 101, pp 271-285

Kuwata, K. and R. Shibasaki (2015) Estimating crop yields with deep learning and remotely sensed data, *2015 IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*, DOI: 10.1109/IGARSS.2015.7325.900

Kuwata, K. and R. Shibasaki (2016) Estimating corn yield in the United States with MODIS EVI and machine learning methods, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume III-8, 2016 XXIII ISPRS Congress*, 12–19 July 2016, Prague, Czech Republic

Li, A., S. Liang, A. Wang and J.Qin (2007) Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques, *Photo. Eng. Remote Sensing*, Vol. 73, No.10, pp 1149-1157

Liang, S., H. Fang, G. Hoogenboom, J. Teasdale and M. Cavigelli (2004) Estimation of crop yield at the regional scale from MODIS observations, Conference Paper, DOI: 10.1109/GRASS.2004.1370640 – source : IEEE Xplore

Lopresti, M. F., C.M. Di Bella and A. J. Degioanni (2015) Relationship between MODIS-NDVI data and wheat yield: a case study in Northern Buenos Aires province, Argentina, *Information Processing in Agriculture*, doi: <http://dx.doi.org/10.1016/j.inpa.2015.06.001>

Mkhabela, M.S., P. Bullock, S. Raj, S. Wang and Y. Yang (2011) Crop yield forecasting on the Canadian prairies using MODIS-NDVI data, *Agri. And Forest Meteo.*, Vol. 151, No. 3, pp 385-393

NRAA (2012), Prioritization of rainfed areas in India, National Rainfed Area Authority, Planning Commission, Govt. of India, Study report No.4, p 76, 2012

Nitze, I, U. Schulthess and H. Asche (2012), Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification, *Proceedings of the 4th GEOBIA*, May 7-9, 2012 - Rio de Janeiro - Brazil. p.035

Pantazi, X.E., D. Moshou, T. Alexandridis, R.L. Whetton and A.M. Mouazen (2016) Wheat yield prediction using machine learning and advanced sensing techniques, *Computer and Elec. in Ag.* Vol. 121, pp 57-65

Peng S, J Huang, JE Sheehy, RC Laza, RM Visperas, X Zhong, GS Centeno, GS Khush, KG Cassman (2004), Rice yields decline with higher night temperature from global warming, *Proc. Natl. Acad. Sci. USA*, 101(27), 971-9975

Prasad, A. K., L. Chai, R. P. Singh and M. Kafatos (2006) Crop yield estimation model for Iowa using remote sensing and surface parameters, *Int. Jour. Appl. Earth Obs.*, 8, 26-33

Rathnayake, W.M.U.K., R.P.De Silva and N. D.K. Dayawansa (2016) Assessment of the suitability of temperature and relative humidity for rice cultivation in rainfed lowland paddy fields in Kurunegala district, *Tropical Agricultural Research*, 27, No.4, pp 370-388

Rudnicki, W. R., M Wrzesien and W Paja (2015), All Relevant Feature Selection Methods and Applications, Chapter 2 of *Feature Selection for Data and Pattern Recognition, Studies in Computational Intelligence* 584, DOI 10.1007/978-3-662-45620-0\_2, Eds. U. Stanczyk and L C Jain, Springer-Verlag,

Welch, J.R. , J.R. Vincent, M. Auffhammer, P. F. Moya, A. Dobermann, and D. Dawe (2010) Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures, *PNAS*, Vo. 107, No. 33, pp 14562-14567

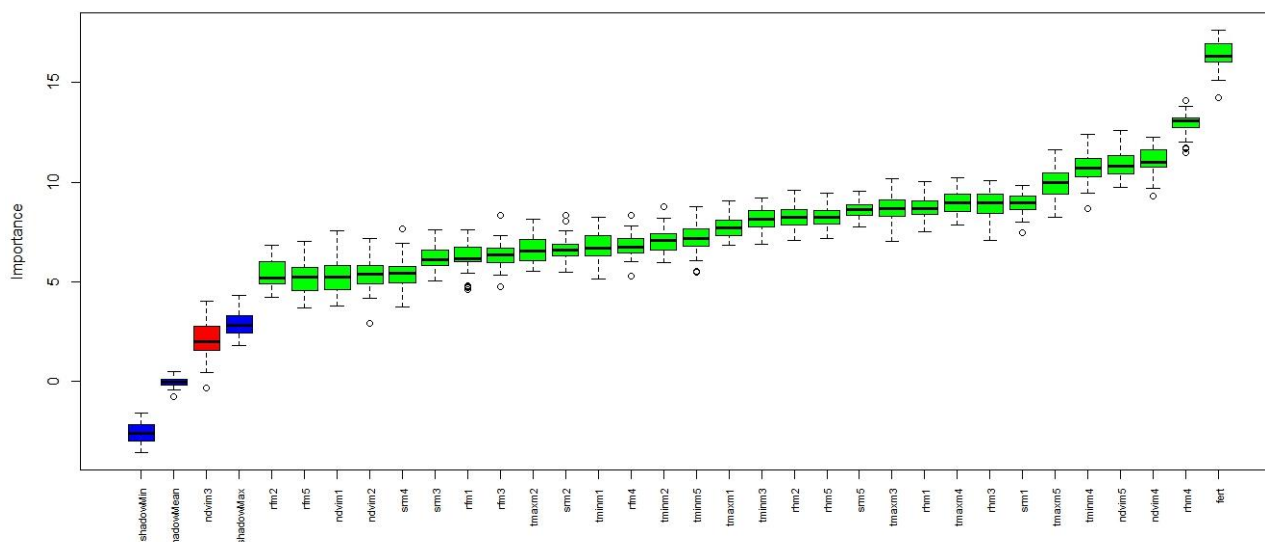
You, J., Li, X., Low, M., Lobell, D., Ermon, S. (2017) Deep Gaussian process for crop yield prediction based on remote sensing data. In: *Proceedings of the 31<sup>st</sup> AAAI Conference on Artificial Intelligence (AAAI-17)*, Deptt. Computer Science, Stanford Univ., U.S.A. 4559-4565

Month	Rainfall (mm)		Temperature (°C)		Solar Radiation (MJ/m <sup>2</sup> /d)		Relative Humidity (%)		NDVI	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
<b>Purulia</b>										
June	51.3	610.4	24.6-28.4	32.4-39.9	12.5	23.8	51.3	80.3	0.266	0.597
July	86.4	736.7	24.1-26.0	30.1-33.5	13.4	19.3	75.1	89.1	0.223	0.613
August	82.3	539.1	24.0-25.3	29.0-32.8	12.6	18.6	79.1	90.5	0.322	0.738
September	42.1	484.6	22.9-24.6	28.6-31.4	13.4	18.6	81.8	91.0	0.564	0.775
October	0.0	429.9	17.1-22.0	27.3-30.4	12.5	19.4	70.8	87.7	0.381	0.748
<b>Bankura</b>										
June	35.0	471.1	25.5-28.4	32.4-39.2	16.3	23.9	56.0	82.0	0.234	0.643
July	115.7	770.9	25.0-26.6	30.5-33.3	13.4	19.3	77.4	88.1	0.25	0.784
August	69.2	514.7	25.0-25.9	30.0-32.8	12.6	19.1	79.7	90.3	0.325	0.756
September	122.3	485.6	24.1-25.2	29.6-31.5	13.5	18.9	83.2	91.1	0.421	0.813
October	7.4	423.5	18.3-23.0	28.3-30.4	13.2	19.2	76.8	87.7	0.3	0.808

**Table 1.** Block level variations in monthly weather and NDVI in Purulia and Bankura districts, West Bengal during 2006-2014 (source: IMD, NASA Power & MNCFC).

Purulia district–blocks yield range (kg/ha) 2006-15				Bankura district–blocks yield range (kg/ha) 2006-14			
Arsha	2287-2846	Manbazar - I	1398-2929	Bankura-I	2144-3176	Kotulpur	2862-3577
Baghmundi	1445-3377	Manbazar -II	1310-3040	Bankura-II	2173-3271	Mejia	2367-3036
Balrampur	1768-2886	Neturia	1731-2630	Barjora	2685-3250	Onda	1713-3111
Banduan	1025-2791	Para	1293-2731	Bishnupur	2386-2908	Patrasayar	2413-3576
Barabazar	1542-2716	Puncha	1449-2876	Chhatna	2201-3255	Raipur	2382-3420
Hura	1606-2664	Purulia - I	1530-2647	Gangajalghati	1290-3386	Ranibandh	2074-3048
Jhalda-I	2017-2823	Purulia - II	1486-2717	Hirbandh	2479-3023	Saltora	1453-2711
Jhalda II	1533-3002	Raghunathpur I	1312-2737	Indpur	1453-2828	Sarenga	2199-2912
Jaypur	1641-2917	Raghunathpur II	1548-2350	Indus	2419-3227	Simlupal	2132-3074
Kashipur	1588-3277	Santuri	985-2397	Joypur	2710-3387	Sonamukhi	2225-3063
				Khatra	2228-3022	Taldangra	2224-3008

**Table 2.** Range of kharif rice yield variations during 2006-2015 in blocks of Purulia & Bankura districts



**Figure 2.** Boruta results box plot for Purulia and Bankura districts monthly NDVI, weather variables and Fertilizer for all relevant feature selection. Blue box plots correspond to minimal, average and maximum Z score of shadow variable. Green box plots represent Z scores of confirmed variables. Red box plot represents rejected variable. The top ranking relevant variables with decreasing rank are fertilizer, RH (Sept), NDVI (Sept), NDVI (Oct), TM (Sept), TX (Oct).

WV – YIELD (2006-2015)		WV WITH FERTILIZER – YIELD.					NDVI-YIELD					NDVI WITH WV - YIELD					NDVI WITH WV, & FERTILIZER - YIELD				
<b>PURULIA (2006-2015) : 9 yrs</b>																					
All	Boruta	RF	All	Boruta	RF	All	2,3,4,5	3,4,5	4,5	All	Boruta	RF-3	RF-7	All	Boruta	RF					
Non-conv.	0.674	0.698	Non-conv.	0.643	0.635	0.534	0.571	0.446	0.431	0.58	0.633	0.668	0.685	Non-conv.	0.616	0.660					
	TX-3,4 3,4,5 RH-3,4 TM-2,4,5 SR-5	TX-3,4 RH-3,4 TM-2,3 SR-5 FERT		TX-3,4,5 RH-3,4 TM-2,4 SR-5 FERT	TX-3,4,5 RH-4 TM-2,4 FERT						TX-3,4,5 TM-2,3,5 RH-3,4 SR-4 NDVI-5	TX-3,4 NDVI-4	TX-3,4,5 RH-3,4 SR-5 NDVI-4		TX-3,4,5 RH-3,4 TM-2,3 SR-5 FERT NDVI-5	TX-3,4 RH-4, NDVI-4,5					
<b>PURULIA (1992-2015) : 24 yrs</b>																					
0.669	0.723	0.702	0.642	0.723	0.702																
	SR-3,4,5 TX-1,5 RH-1,3,5 RH-1	SR-5 TX-1,5 RH-1,3,5 FERT: low rank		SR-3,4,5 TX-1,5 RH-1 FERT: low rank	SR-5 TX-1,5 RH-1,3,5 FERT: low rank																
<b>BANKURA (2006-2014)</b>																					
All	Boruta	RF	All	Boruta	RF	All	2,3,4,5	3,4,5	4,5	All	Boruta	RF	All	Boruta	RF						
0.381	0.515	0.515	Non-conv.	0.478	0.512	0.436	0.440	0.455	0.443	0.432	0.484	0.542		0.403	0.415						
	RH-2,3,4 TM-3 RF-3 TX-3,5	RH-2,3,4 TM-3 RF-3 TX-3,5		RH-2,3,4 TM-3 FERT	RH-2,3,4 TM-3 FERT TX-5						RH-1,2,3,4 TX-2,3,5 TM-3 RF-3 NDVI-4	RH-4 RF-3 TX-5 TM-3 TM-3 NDVI-4		RH-3,4 TM-3 TX-5 RF-3 NDVI-4 FERT	RH-4 TX-5 TM-3 RF-5 NDVI-4 FERT						
<b>PURULIA &amp; BANKURA (2006-2014/15)</b>																					
All	Boruta	RF	All	Boruta	RF	All	2,3,4,5	3,4,5	4,5	All	Boruta	RF	All	Boruta	RF	Boruta					
0.549	0.649	0.642	0.537	0.682	0.659	0.589	0.604	0.565	0.567	0.495	0.651	0.663	0.527	0.684	0.662	0.700					
	RH-1,4 TM-4 TX-5 TM-3,4 SR-1	RH-4 TX-5 TM-3,4		FERT RH-3,4 TX-4,5 TM-4	FERT RH-3,4 TM-3,4 TX-5						RH-4 TM-4 NDVI-4,5 TX-5	RH-4 NDVI-4,5 SR-5 TX-1		FERT RH-3,4 NDVI-5,4 TM-4 TX-4,5	FERT RH-3,4 NDVI-5,4 TX-5	RH-4 FERT NDVI-4 5					
WV: Weather Variables, NDVI: Normalized Difference Vegetation Index, Boruta: All Relevant Feature Selection Algorithm, RF: Random Forest for Predictor Variable Importance, RH: Relative Humidity, SR: Solar Radiation, TX: Max. Temp., TM: Min. Temp., RF: Rainfall, ALL: June-Oct., 2: July, 3: Aug., 4: Sept, 5: Oct., Non-Conv. : Non-Converging, <i>The model average correlation values given in columns are based on 50 model runs</i>																					

**Table 3.** Results of rice yield prediction models developed for different combinations of predictor variables for Purulia and Bankura districts, West Bengal