International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning Based Techniques for Paddy Yield Prediction for the State of Andhra Pradesh

Ms. Sangeetham Rohini [1], Dr. S. Narayana Reddy[2]

Submitted:10/02/2023 **Revised:**12/04/2023 **Accepted:**07/05/2023

Abstract: Timely and accurate crop yield prediction serves as a pillar for the country's food security and frames the strategic policies for the government. In this study, we endeavoured to assess the effectiveness of three various machine learning-based methods to predict paddy yield for the Indian state of Andhra Pradesh. The models were developed using historical yield data for the years 2001 to 2020 along with the long-term derived satellite variables evapotranspiration (ET), leaf area index (LAI), land surface temperature (LST), normalised difference vegetation index (NDVI), and rainfall (RF). Multiple linear regression (MLR), support vector regression (SVR), and random forest regression (RFR) models were three different machine learning models that were assessed for performance. A correlation was established between these variables and crop yield. The highly correlated features model was built and the features with the least correlation were discarded. The performance of all three models was found to be satisfactory. The RFR model was found to have higher accuracy with an R^2 value of 0.61 and an RMSE of 0.55 t ha^{-1}. Whereas MLR and SVR were found to have R^2 0.51 and 0.59, RMSE 0.59 t $ha^{-1,}$ and 0.54 t ha^{-1}. The results from the current study have shown the capability of machine learning algorithms with limited datasets.

Keywords: *- food security, crop yield, machine learning, RFR, correlation*

1. Introduction

The world's most important crop for staple foods is paddy (Oryza sativa L.). Paddy is produced and exported by India in second- and third-place finishes globally. For the years 2020–2021, India will produce 12.27 million metric tonnes of paddy. The Indian economy is heavily dependent on the paddy crop. India's paddy output climbed from 3.6 t ha-1 in 2011–2012 to 4.2 t ha-1 in 2020–21. India has been a major contributor to the global production of Paddy with a share of 21.81% (2015-16) (Manjunatha et al., 2015).

Crop yield is a key factor (Bender & Heijden, 2015) in Paddy production its management is very crucial for making agricultural policies, trading, forecasting, and adopting a strategic plan to tackle climate change (Jeong et al., 2016). The changing environmental factors and extreme weather events have led to variability in the Paddy yield (Ray et al., 2015; Debnath et al., 2021). This event has become a major concerning factor for farmers and governments to reinforce thinking of the need for accurate and precise crop yield forecasting models (Rashid et al., 2021). Global food security has been seriously threatened by an abrupt rise in global warming and population expansion (Lobell & Bruke, 2008; Tilman et al., 2011). According to Frieler et al. (2017), the impact of agronomic and climatic conditions on crop development and production is unpredictable. Agro-meteorological factors have been heavily incorporated into yield prediction models up to this point (Fang et al., 2011; Li et al., 2015).

The two most popular modelling techniques to forecast crop output in response to agro-meteorological factors are process-based models and statistical models (Prasad et al., 2022). Although process-based crop models can accurately predict crop yield at the field level by simulating the physiological processes of crop growth and development, it is challenging to use them for timely predictions on a regional or global scale due to their high data and calibration needs (Lobell & Bruke, 2008). Without taking into account the uncertainty brought on by physiologic and ecological factors, statistical models attempt to map a direct link between crop yield and predictor

[1]Research Scholar [2]Professor
[1,2]Department of ECE, Sri Venkateswara University College of Engineering,
Tirupati, Andhra Pradesh,India
Email ID:srohini512@gmail.com

variables. Inputs for these extensively are dependent on the crop-cutting experiments and observed agro-meteorological parameters which fail to present these variables at a spatial scale.

Crop production projections may now be made with accuracy because to advancements in remote sensing and Geographic Information Systems (GIS) (Setiyono et al., 2014; Huang et al., 2019). Since their direct correlation with biomass production, biophysical variables such as NDVI (Normalised Difference Vegetation Index), Leaf Area Index (LAI), Evapotranspiration (ET), Land Surface Temperature (LST), Fraction of Photosynthetically Active Radiation (FPAR), etc. have been widely used in crop yield prediction models (Doriaswamy et al., 2003; Supit et al., 2012; Sakamoto et al., 2013). Remote sensing-derived variables have shown to be reliable datasets in representing the physiological conditions of crops on a spatial scale basis. Integration of ground data with remote sensing data has quite promising results in crop yield prediction. Traditional statistical models have shortcomings in capturing the uncertainty when predicting agricultural output with the integration of both remote sensing data and observed data (Chaurusiya et al., 2017; Dubey et al., 2018). The classic modelling approaches have several drawbacks, which are partially addressed by machine learning techniques for crop yield prediction (Schumacher et al., 2019).Machine learning techniques have gained wider popularity as crop prediction models for accurate yield predictions in India. We can see many examples where machine learning models are used for yield predictions for the Indian case. A few examples are Jaikala et al., 2008 developed a Paddy yield prediction model using a support vector regression (SVR) algorithm and compared the results with the DSSAT model yield. Integrating Sentinel-1 & 2 data paddy mapping and yield prediction was done using a random forest algorithm in the Sahibganj region (Ranjan & Parida, 2019). Guruprasad et al., 2019 used ML techniques to quantify the paddy yield at district and sub-district levels using weather and soil data as predictors. They have achieved an average error of 3.14% in predicting the paddy yield. Nihar et al., 2022 evaluated different machine learning models to predict sugarcane yield for the state of Uttar Pradesh using long-term satellite-derived variables and crop yield data. They have achieved a 66% of yield prediction accuracy using Gradient Boost Regression (GBR) with an RMSE of 7.15 t ha^{-1}. Arumugam et al., 2021 have successfully downscaled the paddy crop yield at the district level using coarse resolution LAI and crop mask. They have used the GBR algorithm to downscale and re-aggregate the paddy yield for district-level yield predictions. The GBR-modeled yield was found to be in agreement with the observed yield. To develop accurate and precise crop prediction models understanding the efficacy of quality of input parameters plays a very important role.

Finding the optimum crop production prediction model for paddy was the goal of this study. We evaluated the performance of three different machine learning models for predicting paddy yield at a regional scale using long-term derived satellite variables and historical crop yield data. The best model for predicting paddy yield with the lowest percentage of error was determined by comparing these models.

2. Materials and Methods

2.1 Study Area

The main food crop in the Indian state of Andhra Pradesh is paddy. The state is located between latitudes 12°41' and 19.07°N and longitudes 77° and 84°40'E. Telangana borders the state to the northwest, Chhattisgarh to the north, Orissa to the northeast, Tamil Nadu to the south, and Karnataka to the southwest. It is the seventh-largest state in the union and occupies 162,975 km2, or 4.96%, of the nation's total land area (Figure 1). Godavari and Krishna are the two major rivers that run across the state. Andhra Pradesh generally experiences a hot and humid type of climatic condition with a mean annual temperature ranging between 21° and 40°C. Andhra Pradesh's coastal lowlands get hotter summers than the rest of the state. In Andhra Pradesh, the rainy season lasts from July through September. During this time, when the southwest monsoon is in full effect, the state experiences the most rainfall. Additionally, it gets a third of the northeast monsoon's rainfall during that time. The state typically receives 1045 to 1170 mm of rainfall per year (Guhathakurta et al., 2020). The primary and staple crop in Andhra Pradesh is paddy. In the state, it is grown on more than 22 lakh hectares throughout both the Kharif and Rabi seasons. 7 million metric tonnes of rice were produced in the state altogether for the year 2020-21. Cotton, Groundnut, Pigeon pea, Sunflower, Black gram, and Sorghum are the other major crops grown in the state along with the Paddy.

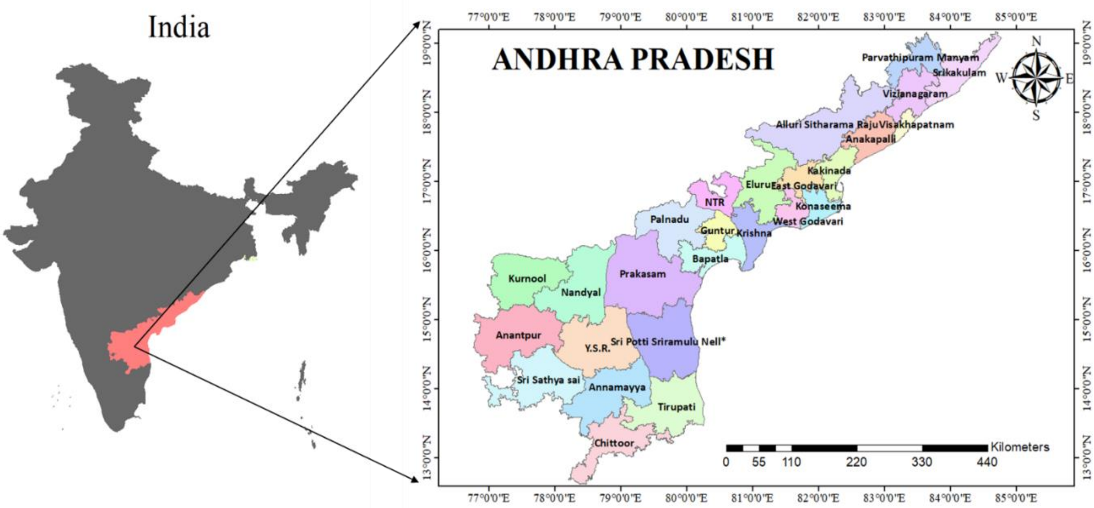


Fig. 1 – Location of Andhra Pradesh State along with its district labels.

2.2 Materials

The datasets utilised in this investigation were obtained from the Terra satellite's MODIS (Moderate Resolution Imaging Spectro Radiometer) instrument. Within two days of its revisit time, the MODIS sensor has a very wide swath of 2,336 km and can cover the whole planet. It offers imagery in 36 multispectral bands with 250 m, 500 m, and 1000 m spatial resolution. According to Patel et al. (2006), it provides ready-to-use, high-value geophysical products that directly address the demand for regional and global level monitoring. Evapotranspiration (ET), Leaf Area Index (LAI), Land Surface Temperature (LST), and Normalised Difference Vegetation Index (NDVI) are the MODIS land products used in this study (Table 1) (Didan, 2021).

Table 1 – Description of satellite datasets used in the study

Satellite/Sensor	Product	Resolution	Frequency
MODIS (Terra)	ET (MOD16A2)	500 meters	2001-2021 (8 days)
	LAI (MOD15A2H)	500 meters	2001-2021 (8 days)
	LST (MOD11A1)	1000 meters	2001-2021 (daily)
	NDVI (MOD13Q1)	250 meters	2001-2021 (16 days)
CHIRPS	Rainfall	0.05°	2001-2021 (daily)

The CHIRPS (Climate Hazards Group InfraRed Precipitation with Station based platform) data for this investigation were obtained from Google Earth Engine (Gorelick et al., 2017), which is openly accessible in the public catalogue.Datasets were converted to monthly averages and the mean value was extracted using the district administrative boundaries of Andhra Pradesh state.

Historical Paddy yield data for Andhra Pradesh state was taken from the Directorate of Economics and Statistics (DES, 2021). The database consists of district-wise area (in hectares), production (tonnes), and productivity information (tonnes per hectare) for all the states of India. By filtering the parameters, the Kharif season Paddy yield data were retrieved for all the districts of Andhra Pradesh. The Paddy yield data obtained was in the unit of “ton per hectare”.

2.3 Methodology

This study is based on a simplistic approach to predicting regional-level Paddy crop yield using machine learning-based algorithms with the aid of historical crop yield data (Paddy) and long-term available satellite variables derived for the study area. From the years 2001 to 2021, the entire state of Andhra Pradesh had its satellite data downloaded (Table 1).. The satellite variables with different temporal and spatial resolutions were zonal averaged to a monthly scale using district administrative boundaries. Since the Kharif season Paddy is grown from June to October month in Andhra Pradesh the average values for five satellite variables were taken only for these months as our annual crop yield data was also for the Kharif season only. The variables were subdivided with the corresponding months (June as 06, July as 07, August as 08, September as 09, and October as 10) ET_06, ET_07, LAI_06 so forming a total of 25 (5×5) variables. The data table structure was organized in such a way that the table had columns of Districts, Year, Yield, and variables arranged to correspond with the months with the codes as seen in example Table 2. Each year has monthly scale satellite variables (independent variables) corresponding with annual Paddy yield data (dependent variable) district-wise from 2001-2020 making a total of 230 observations.

Table 2 Example of datasets organized in table for training ML models

Districts	Year	Yield	ET_06	ET_07……	NDVI_06	NDVI_07……
ANANTAPUR	2001					
CHITTOOR	2001					
EAST GODAVARI	2001					
GUNTUR						
ANANTAPUR	2002					
CHITTOOR	2002					

Regression analysis works as a predictive modelling technique that establishes the relationship between dependent (target) and independent (predictors) variables. Data has to be cleaned for the outliers for better performance of the model. The entire analysis for the yield modelling was carried out in the cloud platform Google Collab python (Bisong, 2019) environment. The data were rescaled using the StandardScaler() function available within the Sci-kit learn package to bring all the datasets in the specific range. Now the data consisted of 25 variables from 5 variables from 2000 to 2020.

The collinearity test between the crop yield and predictor variables was computed. The poorly correlated variables were removed since they didn't have any significance in the construction of models. The data had 228 rows and 28 columns, each row corresponded to each district and the year for which crop yield and satellite variables were extracted. The dataset was split into a 70:30 ratio as training and testing datasets. 70% of datasets were used in the model construction and the remaining 30% of datasets served as validation data. The detailed workflow for the overall methodology can be seen in Figure 2.

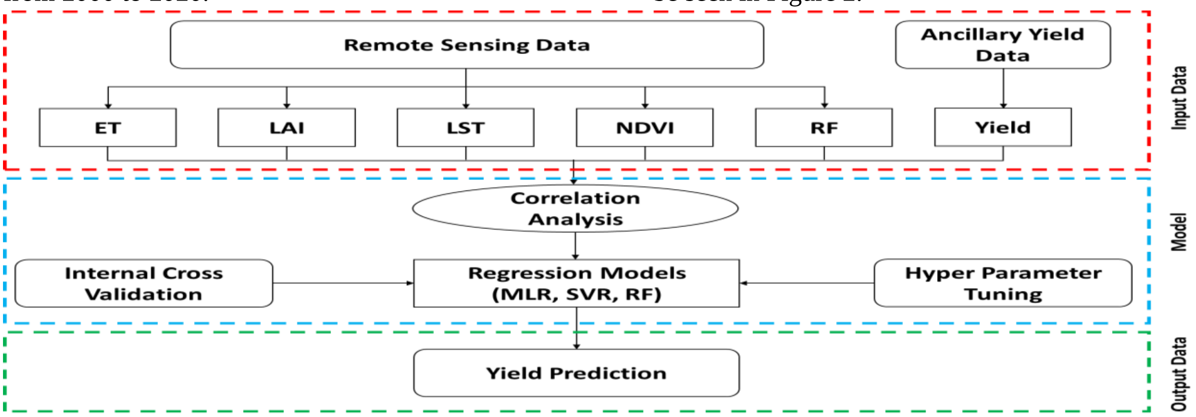


Fig. 2 – Methodology Flow Chart

Three machine learning techniques were employed in this study to forecast the Paddy yield. The advantages and cons of each algorithm are different. Multiple linear regression, support vector regression, and random forest regression were the three machine learning techniques used in this study. MLR makes an effort to model the dependent variable as a linear function of two or more independent variables. SVR, on the other hand, is a form of Support Vector Machine (SVM) used to predict continuous variables. SVR is used to fit a hyperplane that minimises the error between the predicted and actual values rather than searching for a border between distinct classes. Then, fresh data points can be predicted using this line. RFR is a particular kind of ensemble learning approach for regression issues. It consists of many decision trees, each of which was trained using a random subset of the data and the characteristics. The average of all the forest's trees' projections is used to determine the outcome. Powerful and resistant to overfitting, RFR is an algorithm that can handle high-dimensional data.

Pre-processed datasets were trained using the models. Based on the correlation analysis only the significant features were selected using the SelectFromModel() function available within the sci-kit learn library. By minimising the loss function, hyperparameter tuning is used to identify the best set of hyperparameters that produces the highest model performance when applied to a set of independent data.. To estimate the performance, fivefold cross-validation is used. The final model is selected when it reaches a reasonable level of accuracy while also having a reasonable number of parameters.

Based on statistically significant metrics like R^2, Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE), the performance of the models was assessed and contrasted. The best yield prediction model is chosen based on the evaluation of these metrics.

3. Results

In order to choose the most influential variables affecting Paddy's yield a simple Pearson's correlation test was performed. The correlation test between the Paddy yield and the predictor variables is shown in figure 3. Based on the correlation test results and feature importance score the variables with the highest score were chosen to construct the model. The Paddy yield has showed a negative association with the predictor factors NDVI, ET, LAI, and RF. LST, however, has demonstrated a strong relationship with the Paddy yield. The strongest positive link has been found between LST and the months of June and July. ET during June and July has shown the highest negative correlation.

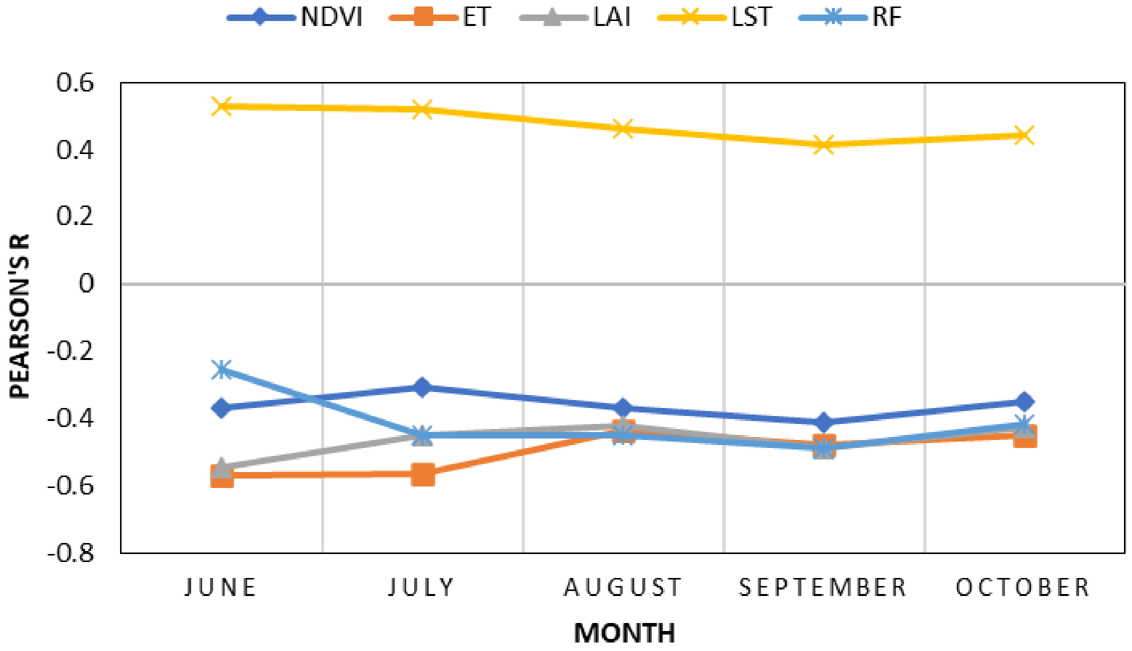


Fig. 3 – Plot showing the relationship between Paddy yield and the monthly mean predictor variables.

The feature priority score was another technique used to choose the variables to be included in the model. The top forecasters were chosen, and the remaining ones were removed, based on the feature importance score. The feature importance score graph for the RFR model is displayed in Figure 4.

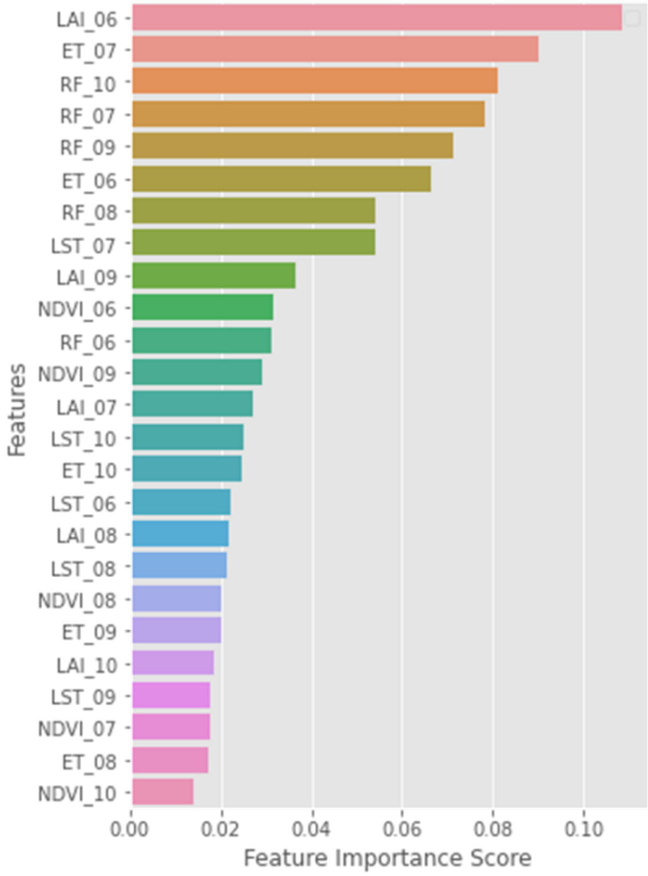


Fig. 4 – Predictor variable importance graph for the RFR model

The top 8 predictor variables were selected out of 25 variables based on feature importance score to construct the model. Fine-tuned hyperparameters were used to construct the models. The performance of the models was evaluated based on the statistical metrics. Out of all three models, the performance of the MLR model was found to be poor when compared with the other two models. RFR performed better than SVR even though both models have almost significant accuracies. The comparison between the models and their statistical signification can be found in Table 3. The results with the highest accuracies have been highlighted in bold.

Table 3 – Evaluation of ML algorithms based on the statistical metrics

Models	R^2	MAE (t ha^{-1})	MSE (t ha^{-1})	RMSE (t ha^{-1})
MLR	0.51	0.45	0.35	0.59
SVR	0.59	0.43	**0.30**	**0.54**
Random Forest	**0.61**	**0.42**	0.31	0.55

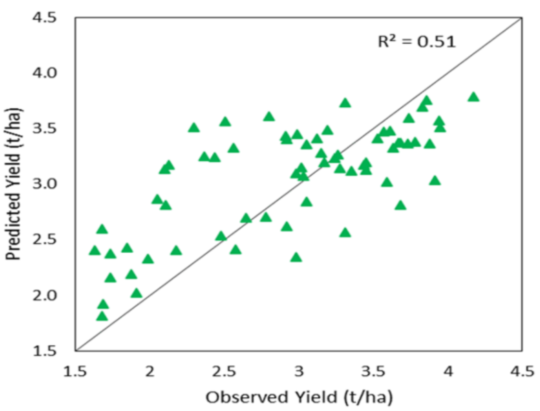
With an MAE of 0.45 t ha-1, MSE of 0.35 t ha-1, RMSE of 0.59 t ha-1, and an R2 value of 0.51 the MLR model was able to forecast the Paddy yield. With an MAE of 0.43 t ha-1, MSE of 0.30 t ha-1, RMSE of 0.54 t ha-1, and R2 value of 0.59, SVR was able to forecast the Paddy yield. While the MAE, MSE, RMSE, and R2 values for the RFR model's prediction of the Paddy yield were each 0.42 t ha-1, 0.31 t ha-1, 0.55 t ha-1, and 0.61 respectively. The Paddy yield that the ML models anticipated differed by area. The maximum recorded yield was found to be 4.33 t ha-1 while the minimum recorded yield was 0.99 t ha-1. The observed average Paddy yield was found to be in the range of 2.9 t ha-1 with a standard deviation of 0.7 t ha-1.

4. Discussion

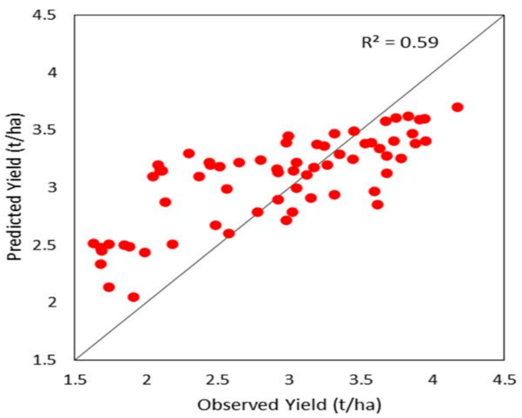
Paddy yield at the regional level for the state of Andhra Pradesh was predicted using long-term historical crop yield data and satellite variables. Among all the predictor variables ET, LAI and LST were found to be the top predictor variables. All three ML models have tended to show moderate accuracy in predicting the Paddy yield.

The correlation between the Paddy yield and the variables such NDVI, ET, LAI, and RF has shown decorrelation or negative correlation (Figure 3). Whereas it has shown a positive correlation with the LST. Generally, we can see a positive correlation between the crop yield and NDVI, ET, LAI, and RF with other crops (Nihar et al., 2022; Prasad et al., 2021b). LST has shown a negative correlation with other crops (Prasad et al., 2021a). Whereas in Paddy crop this was the opposite the variables NDVI, ET, LAI, and RF did not find any positive trend, instead only variable that was positively correlated was LST. A similar result of the decorrelation between Paddy yield and rainfall has been reported by (Sarma et al., 2008; Sandhu et al., 2021). Detrending of Paddy yield with NDVI was also reported by (Huang et al., 2013).

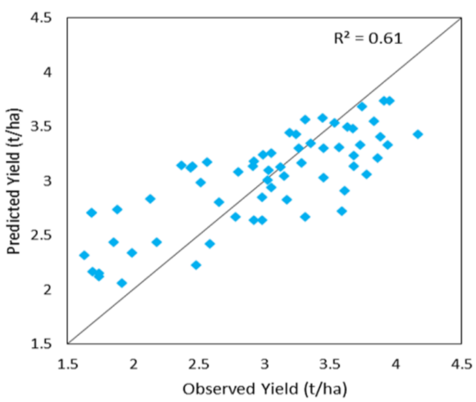
LAI and ET were the top most influential variables influencing the Paddy yield. RF during the starting months had its most effective and LST has its most effective during the second month (during July month) of the crop growth period. Both LAI and ET are the biophysical parameters that affect the photo synthetical activity of the crops. The higher the leaf surface area greater the absorption of radiation, thus LAI is directly proportional to the increase in biomass production. ET is related to the amount of water loss during photosynthesis. Since Paddy is a semi-aquatic crop it requires stagnant water for its growth, thus ET is influencing the yield production of Paddy. LST has also an influence on yield due to the reason that Paddy requires hot humid conditions for its growth. RF from July to October made it to the list of top 8 predictor variables. The southwest monsoon, when Andhra Pradesh receives the majority of its rainfall, is active during these months. NDVI was found to have the least contribution to the yield predicting our top 8 predictor variables.



(a)



(b)



(c)

Fig. 5 – Observed v/s Predicted Paddy Yield for holdout datasets (a) MLR model (b) SVR model (c) RFR model

All three machine learning algorithms predicted the crop yield with satisfactory accuracy. MLR was able to explain the Paddy yield with a 51% (Figure 5(a)) accuracy when tested against 30% independent holdout data that was not used in the training of the model. The SVR model also performed well and was able to explain yield variance with a variability of 59% (Figure 5(b)) accuracy when tested against the holdout datasets. Whereas the RFR algorithm outperformed both MLR and SVM prediction with a 61% (Figure 5(c)) accuracy against the holdout datasets. With the comparative results (Table 3) among all the three machine learning algorithms RFR was found to be most suitable for predicting Paddy yield. SVR performance was also found to be on par with that of RFR results as it had lower values in terms of MAE and RMSE. The results from this study have showcased the potential of RFR algorithms to predict crop yield with higher accuracy. A similar result with higher accuracy for global and regional yield prediction over MLR using RFR has been reported by (Jeong et al., 2016). (Prasad et al., 2021a) has successfully predicted cotton yield in advance for three time periods. Many studies have reported that higher accuracy can be achieved in predicting crop yield using RFR algorithms (Everingham et al., 2016; Geetha et al., 2020; Kumar et al., 2020; Elavarasan et al., 2021; Cheng et al., 2022). The performance of the models tends to be satisfactory but the accuracy of these models has a scope for improvement. The accuracy of these models is mainly confined to the quality of the datasets. Datasets derived in this study were from the MODIS sensor which is a having a coarse resolution (Prasad et al., 2021b). Individual farm size holdings in the Indian context are less than 1 ha which requires higher-resolution satellite datasets to capture the variations in reflectance and avoid mixing of pixels for other field crops' effects. Another drawback for lower accuracy in the results is due to the volume of observed yield data to train the model. The yield data used in this work was at the district-level size, where there were not many examples to train the model. By employing yield data at the block-level scale, the data volume and model accuracy are increased. Only satellite-derived factors were taken into account in this study to establish a correlation with the Paddy yield. But other factors need to be considered other than satellite variables such as fertilizer application, soil type, and other meteorological parameters. The combination of these variables with the satellite-derived variables increases the yield prediction accuracy of these models. Feature study has to be carried out using more advanced algorithms like deep learning algorithms to train the dataset with few observations or instances which has much more promising results than using ML algorithms alone (Dang et al., 2021).

5. Conclusion

Advancement in geospatial technology has made it very much possible to monitor the crops over larger domains and also predict the crop yield before actual harvest. In this study, we have tried to evaluate three different ML algorithms to predict the Paddy yield for the state of Andhra Pradesh. The models were built using historical crop production data together with satellite-derived variables like ET, LAI, LST, NDVI, and RF. The influence of these variables on crop yield was correlated with the crop growing season. The maximum correlation of these variables with crop yield was found to be 35 to 55% with the crop yield. Regression analysis was carried out using MLR, SVR, and RFR models. The RFR model, which had an accuracy of 61% and an RMSE of 0.59 t ha-1, was discovered to be the best predictor model for Paddy yield prediction. When combined with auxiliary yield datasets, this research has demonstrated the usefulness of remote sensing datasets for crop yield prediction. The accuracy of these models has the further scope of improvisation by taking the crop yield data at the block levels and using high-resolution satellite images.

Acknowledgment

The authors are very much thankful to the Google Earth Engine Team for helping the MODIS and CHIRPS datasets from their catalogue. The authors also express sincere thanks to the Ministry of Economics and Statistics for providing the Paddy yield data for the state of Andhra Pradesh.

Declaration

The authors express no conflict of interest.

References

[1] Arumugam, P., Chemura, A., Schauberger, B., & Gornott, C. (2021). Remote Sensing based yield estimation of rice (Oryza sativa L.) using gradient boosted regression in India. Remote

Sensing, 13(12), 2379. https://doi.org/10.3390/rs13122379

[2] Bender, S. F., & van der Heijden, M. G. (2015). Soil biota enhance agricultural sustainability by improving crop yield, nutrient uptake and reducing nitrogen leaching losses. Journal of Applied Ecology, 52(1), 228-239. https://doi.org/10.1111/1365-2664.12351

[3] Bisong, E. Google Collaboratory (2019). Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA, 59-64. https://doi.org/10.1007/978-1-4842-4470-8_7

[4] Chaurasiya, G., Saxena, S., Tripathy, R., Chaudhary, K. N., & Ray, S. S. (2017). Semi physical approach for sugarcane yield modelling with remotely sensed inputs. Vayumandal, 43(1), 11-22.

[5] Cheng, M., Jiao, X., Shi, L., Penuelas, J., Kumar, L., Nie, C., ... & Jin, X. (2022). High-resolution crop yield and water productivity dataset generated using random forest and remote sensing. Scientific Data, 9(1), 1-13. https://doi.org/10.1038/s41597-022-01761-0

[6] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69. https://doi.org/10.1016/j.compag.2018.05.012

[7] Dang, C., Liu, Y., Yue, H., Qian, J., & Zhu, R. (2021). Autumn crop yield prediction using data-driven approaches: -support vector machines, random forest, and deep neural network methods. Canadian Journal of Remote Sensing, 47(2), 162-181. https://doi.org/10.1080/07038992.2020.1833186

[8] Debnath, S., Mishra, A., Mailapalli, D. R., Raghuwanshi, N. S., & Sridhar, V. (2021). Assessment of rice yield gap under a changing climate in India. Journal of Water and Climate Change, 12(4), 1245-1267. https://doi.org/10.2166/wcc.2020.086

[9] DES. (2021). Directorate of economics and statistics, ministry of agriculture and farmers welfare, government of India. https://eands.dacnet.nic.in/

[10] Didan, K. (2021). MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2022-12-23 from https://doi.org/10.5067/MODIS/MOD13Q1.061

[11] Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment from remote sensing. Photogrammetric engineering & remote sensing, 69(6), 665-674. https://doi.org/10.14358/PERS.69.6.665

[12] Dubey, S. K., Gavli, A. S., Yadav, S. K., Sehgal, S., & Ray, S. S. (2018). Remote sensing-based yield forecasting for sugarcane (Saccharum officinarum L.) crop in India. Journal of the Indian Society of Remote Sensing, 46(11), 1823-1833. https://doi.org/10.1007/s12524-018-0839-2

[13] Elavarasan, D., & Vincent, P. M. (2021). A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. Journal of Ambient Intelligence and Humanized Computing, 12(11), 10009-10022. https://doi.org/10.1007/s12652-020-02752-y

[14] Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for sustainable development, 36(2), 1-9. https://doi.org/10.1007/s13593-016-0364-z

[15] Fang, H., Liang, S., & Hoogenboom, G. (2011). Integration of MODIS LAI and vegetation index products with the CSM–CERES–Maize model for corn yield estimation. International Journal of Remote Sensing, 32(4), 1039-1065. https://doi.org/10.1080/01431160903505310

[16] Frieler, K., Schauberger, B., Arneth, A., Balkovič, J., Chryssanthacopoulos, J., Deryng, D., ... & Levermann, A. (2017). Understanding the weather signal in national crop-yield variability. Earth's future, 5(6), 605-616. https://doi.org/10.1002/2016EF000525

[17] Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen (2015). "The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes". Scientific Data 2, 150066. doi:10.1038/sdata.2015.66.

[18] Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, A. P.

(2020, July). An effective crop prediction using random forest algorithm. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE. https://doi.org/10.1109/ICSCAN49426.2020.9262311

[19] Guhathakurta, P., Sanap, S., Menon, P., Prasad, A. K., Sangwan, N., & Advani, S. C. (2020). Observed rainfall variability and changes over Andhra Pradesh state. India Meteorological Department, Pune, India. https://imdpune.gov.in/Reports/rainfall%20variability%20page/reports/andhra_final.pdf

[20] 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, 202, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

[21] Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (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

[22] Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., ... & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and Forest Meteorology, 276, 107609. https://doi.org/10.1016/j.agrformet.2019.06.008

[23] Jaikla, R., Auephanwiriyakul, S., & Jintrawet, A. (2008, May). Rice yield prediction using a support vector regression method. In 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (Vol. 1, pp. 29-32). IEEE. https://doi.org/10.1109/ECTICON.2008.4600365

[24] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. PloS one, 11(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

[25] Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020, June). Supervised machine learning approach for crop yield prediction in agriculture sector. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 736-741). IEEE. https://doi.org/10.1109/ICCES48766.2020.9137868

[26] Li, Z., Wang, J., Xu, X., Zhao, C., Jin, X., Yang, G., & Feng, H. (2015). Assimilation of two variables derived from hyperspectral data into the DSSAT-CERES model for grain yield and quality estimation. Remote Sensing, 7(9), 12400-12418. https://doi.org/10.3390/rs70912400

[27] Lobell, D. B., & Burke, M. B. (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. Environmental Research Letters, 3(3), 034007. https://doi.org/10.1088/1748-9326/3/3/034007

[28] Manjunatha, A. V., & Kumar, P. (2015). Impact of national food security mission (NFSM) on input use, production, yield and income in Karnataka. Agricultural Development and Rural Transformation Centre Institute for Social and Economic Change, 79-81.

[29] Nihar, A., Patel, N. R., & Danodia, A. (2022). Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. Journal of the Indian Society of Remote Sensing, 1-12. https://doi.org/10.1007/s12524-022-01549-0

[30] Patel, N. R., Mohammed, A. J., & Rakhesh, D. (2006). Modelling of Wheat Yields Using Multi-Temporal Terra/MODIS Satellite Data. Geocarto International, 21(1), 43-50. https://doi.org/10.1080/10106040608542373

[31] Prasad, N. R., Patel, N. R., & Danodia, A. (2021). Crop yield prediction in cotton for regional level using random forest approach. Spatial Information Research, 29(2), 195-206. https://doi.org/10.1007/s41324-020-00346-6

[32] Prasad, N. R., Patel, N. R., & Danodia, A. (2021). Cotton Yield Estimation Using Phenological Metrics Derived from Long-Term MODIS Data. Journal of the Indian Society of Remote Sensing, 49(11), 2597-2610. https://doi.org/10.1007/s12524-021-01414-6

[33] Prasad, N. R., Patel, N. R., Danodia, A., & Manjunath, K. R. (2022). Comparative

performance of semi-empirical based remote sensing and crop simulation model for cotton yield prediction. Modelling Earth Systems and Environment, 8(2), 1733-1747. https://doi.org/10.1007/s40808-021-01180-x
[34] Ranjan, A. K., & Parida, B. R. (2019). Paddy acreage mapping and yield prediction using sentinel-based optical and SAR data in Sahibganj district, Jharkhand (India). Spatial Information Research, 27(4), 399-410. https://doi.org/10.1007/s41324-019-00246-4
[35] Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, 63406-63439. https://doi.org/10.1109/ACCESS.2021.3075159
[36] Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. Nature communications, 6(1), 1-9. v https://doi.org/10.1038/ncomms6989
[37] Sakamoto, T., Gitelson, A. A., & Arkebauer, T. J. (2013). MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment, 131, 215-231. https://doi.org/10.1016/j.rse.2012.12.017
[38] Sandhu, S. S., Dhillon, B. S., & Singh, H. (2021). Rice yield variability in Punjab: an overview of five decades. Paddy and Water Environment, 19(4), 673-681. https://doi.org/10.1007/s10333-021-00866-3
[39] Sarma, A. A. L. N., Kumar, T. L., & Koteswararao, K. (2008). Development of an agroclimatic model for the estimation of rice yield. J. Ind. Geophys. Union, 12(2), 89-96.
[40] Schumacher, D. L., Keune, J., Van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Teuling, A. J., & Miralles, D. G. (2019). Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. Nature Geoscience, 12(9), 712-717. https://doi.org/10.1038/s41561-019-0431-6
[41] Setiyono, T., Nelson, A., & Holecz, F. (2014). Remote sensing-based crop yield monitoring and forecasting. Crop monitoring for improved food security.
[42] Supit, I., Van Diepen, C. A., De Wit, A. J. W., Wolf, J., Kabat, P., Baruth, B., & Ludwig, F. (2012). Assessing climate change effects on European crop yields using the Crop Growth Monitoring System and a weather generator. Agricultural and Forest Meteorology, 164, 96-111.
[43] Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the national academy of sciences, 108(50), 20260-20264. https://doi.org/10.1073/pnas.1116437108