AI/ML BASED MODEL FOR CROP MONITORING AND CROP YIELD PREDICTION

Lab Based Project (B. Tech)

04-05-2023

Presented by
Murthati Mahibabu(20114058)
Priya(20114077)
Nimmagadda vasavi(20114064)

Under the supervision of
Prof.Dharmendra Singh
Department of Computer Science and Engineering

Outlines

- ★ Introduction
  - Theoretical background
  - Objectives
  - Motivation
- ★ Literature review/research gaps
- ★ Random Forest and LSTM
- ★ Dataset Analysis
- ★ Flowchart
- ★ Methodology
- ★ Results
- ★ Concluding remarks
- ★ Limitations
- ★ future scope
- ★ References
- ★ Contributions of team members

Introduction

- It is estimated that between 2019 and 2021, around 56 crore Indians, or 40.6 percent of the total population, had moderate or severe food insecurity or inadequate food supply.

- In India the population mostly lives on wheat and paddy,so the predictions of yield for these crops is necessary to avoid inadequate food crises and also important for the farmers to produce more quantitatively.

- We use Satellite remote sensing data (Sentinel-2) as it provides strong advantages over other monitoring techniques by providing a timely, synoptic, and up-to-date overview of large-scale crop monitoring at multiple stages are used to predict the crop yield.

THEORETICAL BACKGROUND

1. The recent rapid development in artificial intelligence has led to an increase machine learning (ML) and deep learning (DL) algorithms, which have been successfully applied in various domains and obviously outperform other traditional techniques.
2. Therefore, some recent studies have applied DL and ML for yield estimation including support vector machines (SVMs), deep neural networks (DNNs), convolutional neural networks (CNNs), and long-short-term memory (LSTM).
3. In particular, the potential of application of available datasets from satellite images and different ML and DL methods for crop yield estimation can provide better results than any other traditional methods.

OBJECTIVES

1.Comparing RF and LSTM algorithms in estimating wheat yield using multi-source data .

2.understanding the major factors that are essential for crop yield estimation by doing prediction analysis for each feature in the dataset.

3.To provide a scalable,simple and inexpensive operational model framework for accurately and timely estimating crop yield.

MOTIVATION

Predicting crop yields is one of the most challenging tasks in agriculture. It plays an important role in decision making at global,regional and filed levels.

In this model we are trying to find an appropriate AI/ML model for crop yield prediction by using timely input data for various features which are essential for crop yield prediction.

As a CS student experimenting on various Ml models and also extracting data for satellite images are also motivating factors for this project.

LITERATURE REVIEW

Model type	Literature review from previous research papers
statistical models	Statistical crop models were fitted with different sample sizes, and it was found that these models are generally able to provide good estimates of impacts of changes in mean and variability of temperature and precipitation when they are fitted based on an adequate sample size.
ML/AI models	Based on different research papers, selected publications use a variety of features, depending on the scope of the research and the availability of data. Every paper investigates yield prediction with machine learning but differs from the features. The studies also differ in scale, geological position, and crop. The choice of features is dependent on the availability of the dataset. Studies also stated that models with more features did not always provide the best performance for the yield prediction. To find the best performing model, models with more and fewer features should be tested. Many algorithms have been used in different studies. The results show that no specific conclusion can be drawn as to what the best model is, but they clearly show that some machine learning models are used more than the others.

LITERATURE REVIEW

Model type	Literature review from previous research papers
Remote sensing data models	Satellite data for crop yield prediction, reviewing traditional and new methods of extracting relevant information from satellite imagery.Satellite images were used to predict farm-scale crop yields, and the results show that models using satellite images improves the prediction accuracy over the baseline model using weather data.Accurate field boundaries, represented using pixel masks and satellite images, significantly improve prediction accuracy using the best performing model.Prediction accuracy was consistently better with models that incorporate both weather data and satellite images, suggesting that both contain some information that the other does not.

Simulation modelling/Algorithms

Model Architecture:

The ML models which we have employed for the Crop yield prediction are – Random Forest, and LSTM(Long-Short-term memory).For tuning the models, the dataset has been divided into training and testing in 47% and 53% in random forest and LSTM model, 2018 and 2019 data has been used for training and 2020 data for testing.

Random Forest Regression Model:

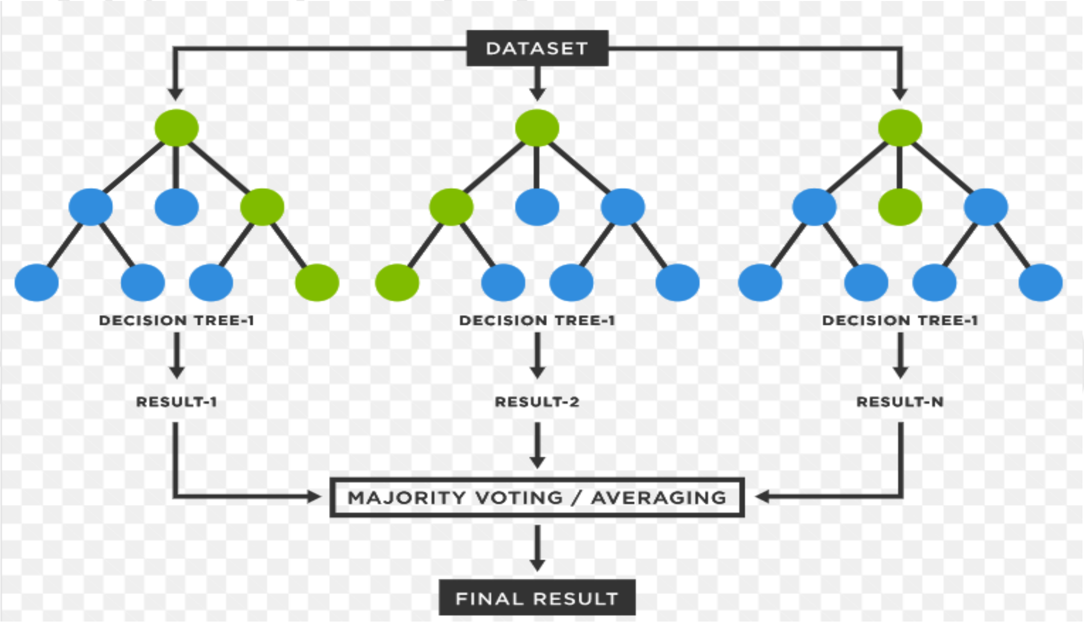
- Random Forest Regression is a popular Machine learning algorithm that can be used crop yield prediction and crop monitoring due to its ability to handle complex and non-linear relationships between the input and output variables.
- It randomly selects a subset of the training data and a subset of the input features to build each decision tree. This process is repeated to build multiple decision trees, each with different subsets of the data and features.
- Outputs the predictions by combining decisions from a sequence of base models in the case of regression problems.Moreover, the RF model is an additive model; more formally, we can express this model as follows:

$$g(x) = f(x) + f1(x) + f2(x) + ... + fn(x)$$

where the model g(x) is the sum of simple base models fn(x). Each f (x) is a simple decision tree. All the base models are independently built using a different subsample of the data.

Random Forest Model:

Basically we will train the model with input data which has features like NDVI,SAVI,DVI,RVI,IPVI,MSAVI and we will try to predict the Yield of a crop.Here, we will try to analyze the each feature of a input by a Random Forest Model and also we will try to analyze the each feature by plotting graph and rmse values.We also calculated yield and rmse value using all the features.



LSTM model:

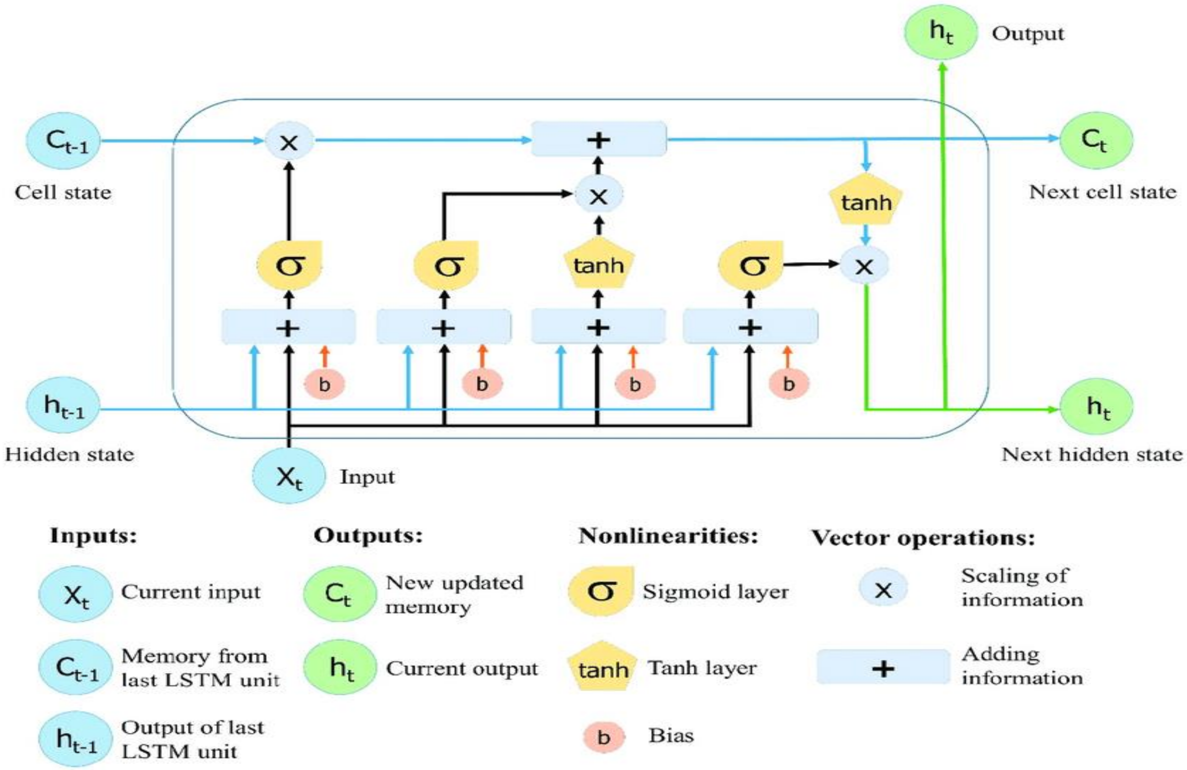
- An LSTM, a basic form of RNN (Recursive neural network), needs to input sequential data.Each time step (ht) depends on the previous step (ht-1) and outside input (xt), then produces output (ot) at this moment and offers this time step (ht) for the next step. Finally, a fully connected layer maps the output.

- The sequential data were dealt with using LSTMs. Finally, a fully connected layer maps the output of the last time step cell to the output node. This LSTM model has 4 time steps and three hidden layers, and each LSTM has 50 hidden units and one Dense layer

- The parameters like epochs, mini-batch size, activation function, dropout, learning rate and optimizer were adjusted according to the model validation requirements.

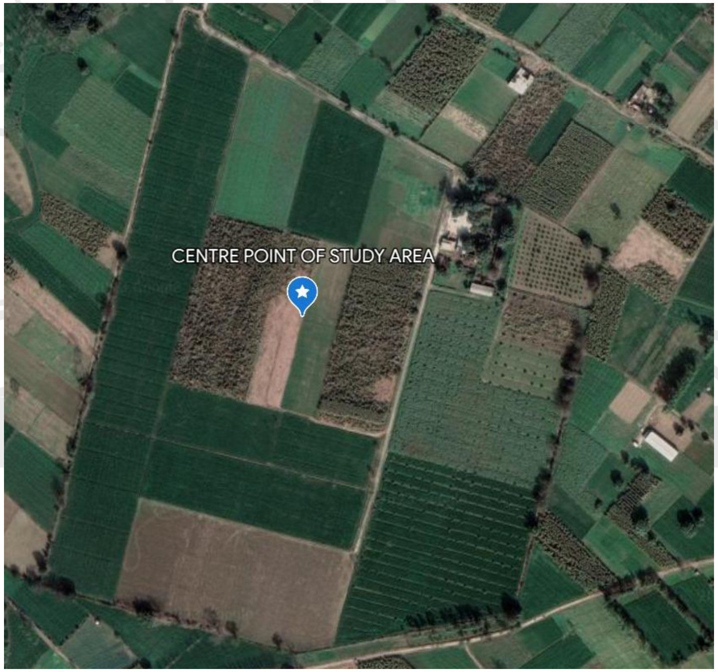
LSTM ALGORITHM:

Reshape this 2D-data into 3D-data for lstm model to capture the timely data properly.Split the data into test data and train data.Train the model with training data and test the model with test data.



Study Area and Dataset

- **Study area:**
  In this project, our center of study for the input data to test the model are mainly from the fields of Jamalpur village in Uttarakhand, Haridwar district. Latitude and longitude for the region of the study are 29°55′50°N,77°57′52°E respectively. The climate here is pleasant, gentle breeze and hot and humid during the months of May and April.



Dataset Analysis

- **Dataset features and remote sensing data:**
  Sentinel-2 is designed to provide high-resolution optical imagery of the Earth's surface

  Sentinel-2 is captured by two identical satellites,Sentinel-2a and sentinel-2b, that orbit the earth at an altitude of 768km. the satellite have a revisit time of five days at the equator, which means that they can capture images of the same location on earth every five days
  The sensors on board the sentinel-2 satellites capture data in 13 spectral bands, ranging from the visible to SWIS(short wave infrared).

  We collected the Remote sensing data from sentinel-2 satellite. The values of NDVI,SAVI,DVI,RVI,IPVI, MSAVI,GNDVI are calculated with the help of this sentinel-2 satellite images.

Dataset Analysis

- **Dataset features and remote sensing data:**

Features	Formulae	Ranges
NDVI(Normalized Difference Vegetation Index)	NDVI = (NIR-Red)/(NIR+Red)	-1 to +1
SAVI(Soil-Adjusted Vegetation Index)	SAVI= (1+L)(NIR-Red)/(NIR+Red+L)	-1 to +1
DVI(Difference Vegetation Index)	DVI = NIR-Red	any value
RVI(Ratio Vegetation Index)	RVI = NIR/Red	0 to infinity
IPVI(Infrared Percentage Vegetation Index)	IPVI = (NIR-Blue)/(NIR+Blue)	-1 to +1
MSAVI(Modified Soil-Adjusted Index.)	MSAVI= (2*NIR+1-sqrt((2*NIR+1)^2-8*(NIR-Red)))/2	-1 to +1
GNDVI(Green Normalized Difference Vegetation Index)	GNDVI = (NIR-Green)/(NIR+Green)	-1 to +1

Dataset Analysis

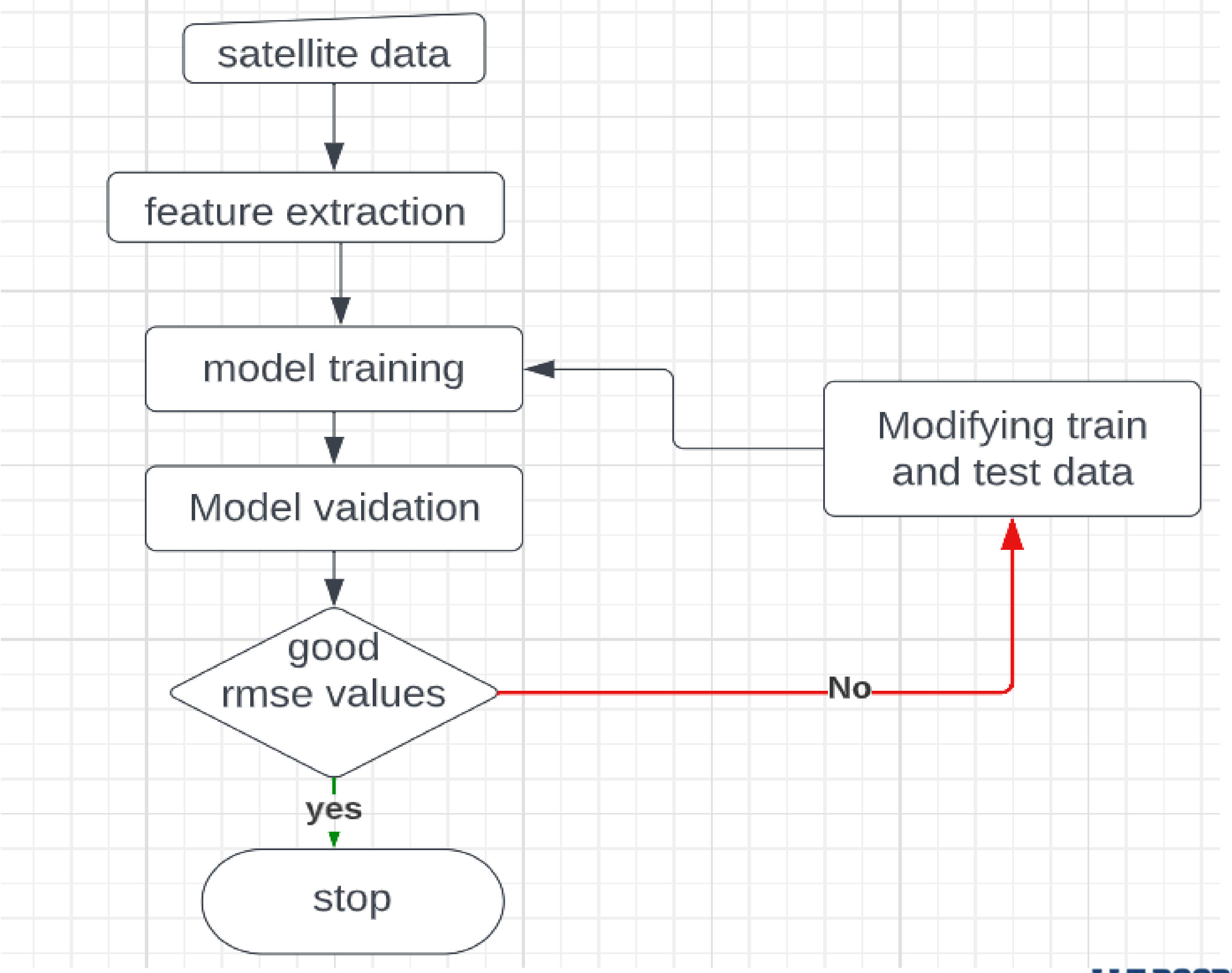
The data here contains the feature values of various fields for four months from 2018-2020.

Here the highlighted part is the data for one field.

Here the output yield is calculated in Quintal/acre.

DATE	NDVI	SAVI	DVI	RVI	IPVI	MSAVI	GNDVI	RAINFALL	YIELD
05-01-18	0.354	0.195	0.103	2.68	0.717	0.218	0.46	1.4	40.1
09-02-18	0.622	0.39	0.117	4.68	0.808	0.436	0.613	1.4	40.1
09-03-18	0.762	0.427	0.164	5.319	0.904	0.513	0.736	13.6	40.1
13-04-18	0.272	0.149	0.087	2.297	0.621	0.126	0.303	42.9	40.1

FLOW CHART



satellite data
feature extraction
model training
Model vaidation
good
rmse values
No
Modifying train
and test data
yes
stop

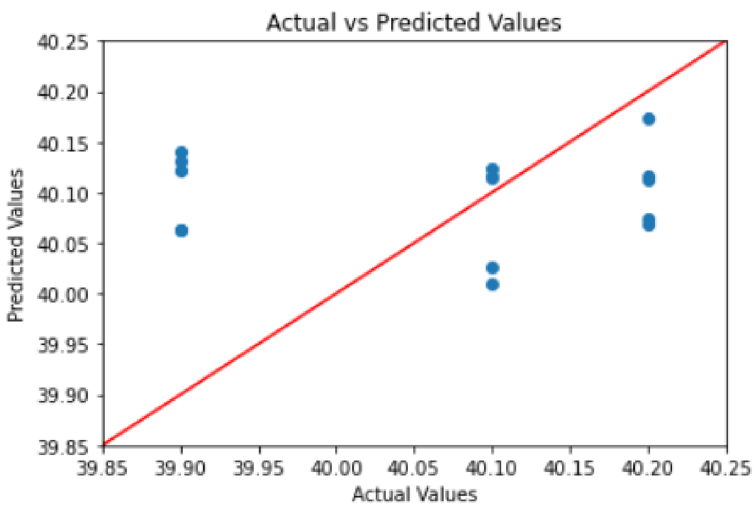
METHODOLOGY

We first collected data values of(NDVI,SAVI,DVI,RVI,IPVI,MSAVI)

from satellite data and we use this data to train the model and perform predictions if we are able to get good rmse values then

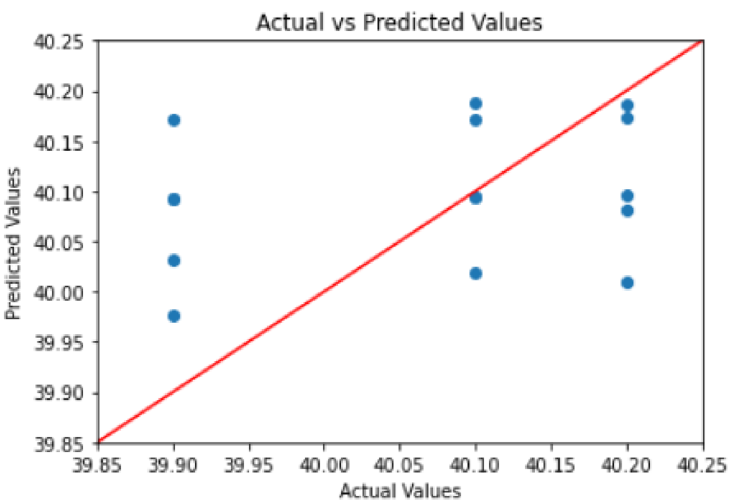
we would stop else we will try to modify the split of test and train data or by modifying hyperparameters values and once again perform predictions to get better rmse values. still if we are not able to get good rmse values then we will look for other models.

RANDOM FOREST RESULTS



MSE: 0.018424982870370633
RMSE: 0.13573865650716685

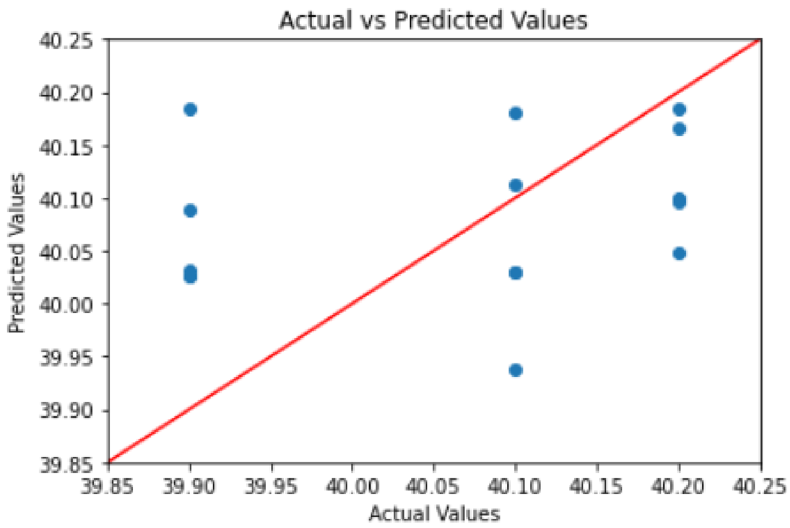
NDVI



MSE: 0.016845795833333618
RMSE: 0.12979135500230213

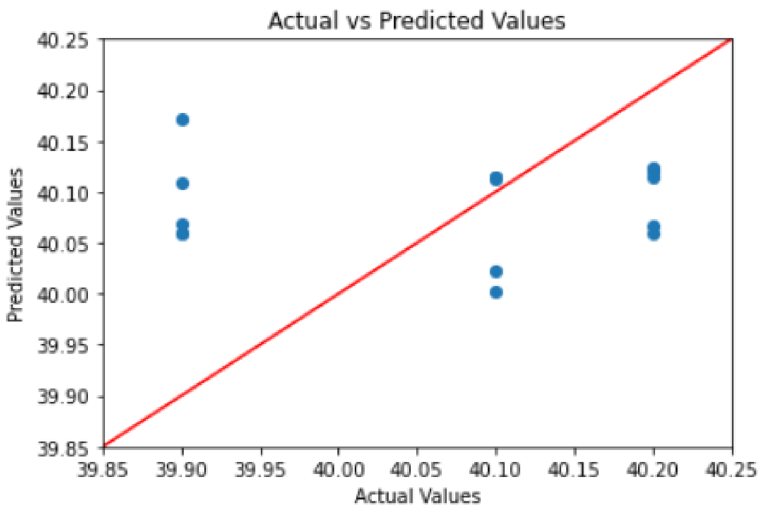
SAVI

RANDOM FOREST RESULTS




```
MSE:  0.01689636215277789
RMSE:  0.12998600752687917
```

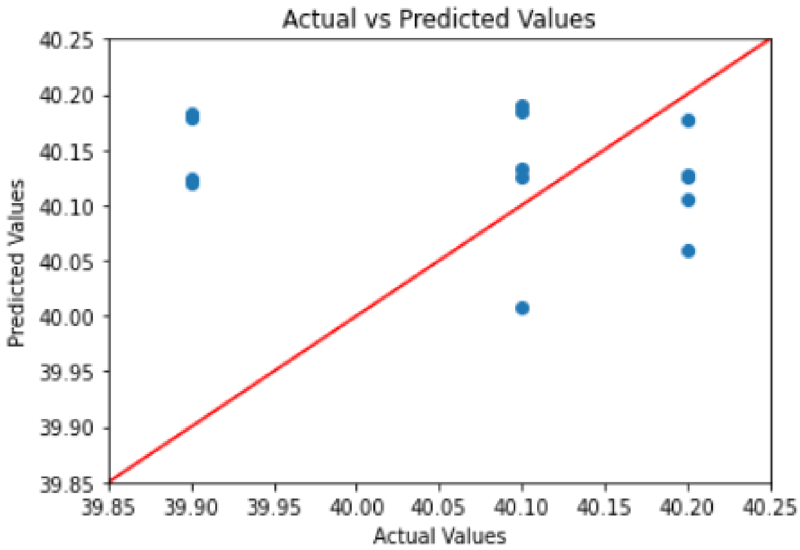
DVI




```
MSE:  0.017953729976852364
RMSE:  0.13399152949665274
```

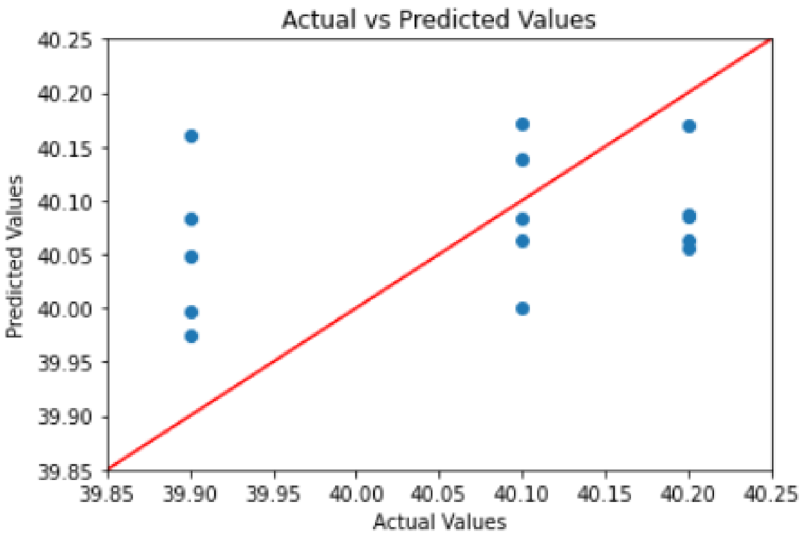
RVI

RANDOM FOREST RESULTS




```
MSE:  0.026664895254630002
RMSE:  0.16329389227595134
```

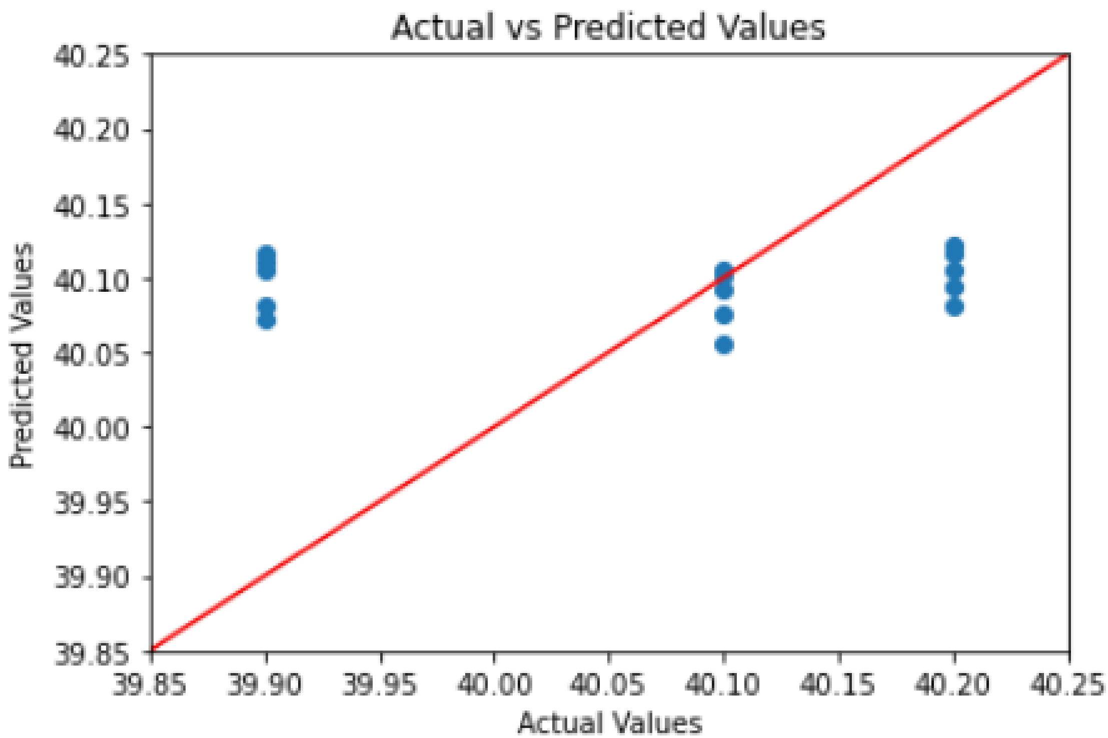
IPVI




```
MSE:  0.014897403703704381
RMSE:  0.12205492085001891
```

MSAVI

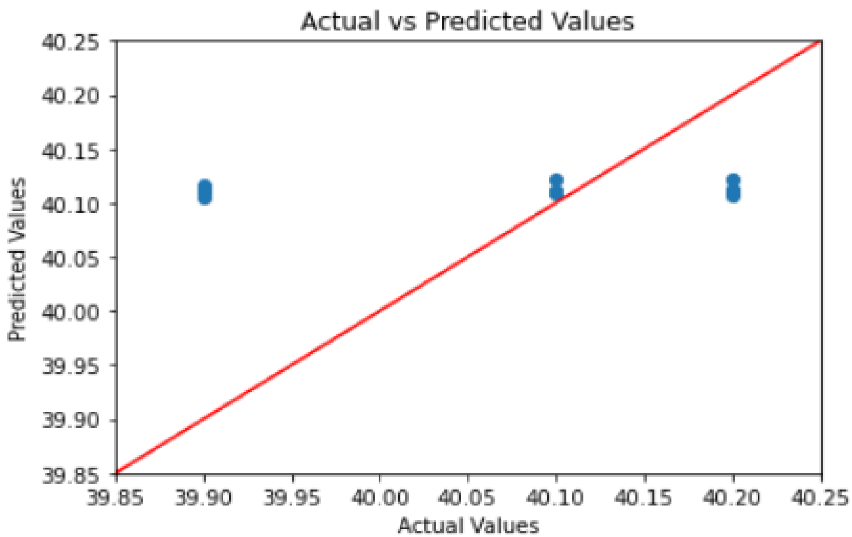
RANDOM FOREST




```
MSE:  0.016404194212963474
RMSE:  0.1280788593522111
```

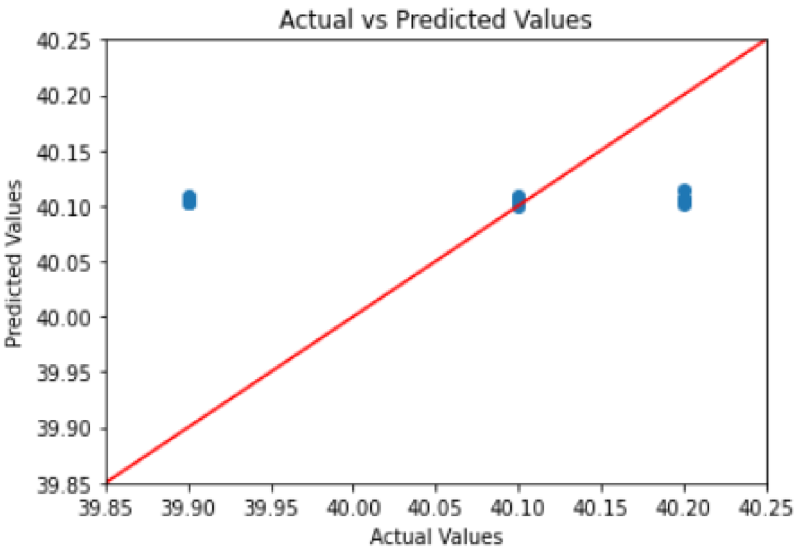
Random forest result by including all features

LSTM RESULTS




```
MSE:  0.017537896797100126
RMSE:  0.1324307245207853
```

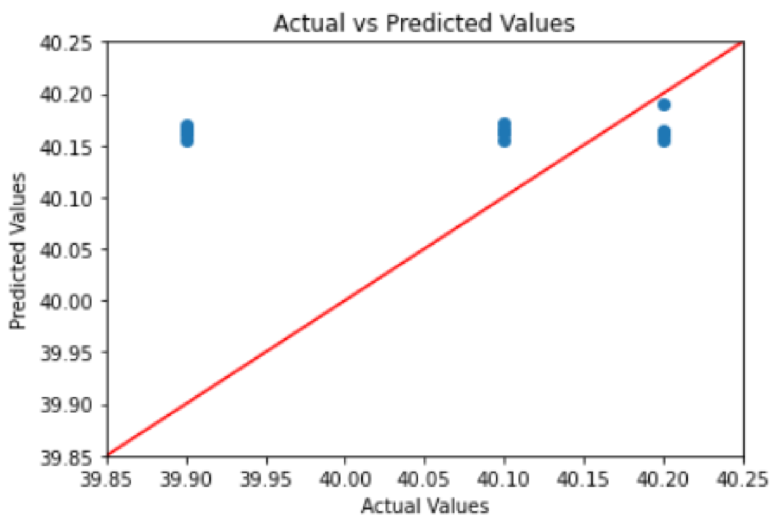
NDVI




```
MSE:  0.01693624916738712
RMSE:  0.13013934519347758
```

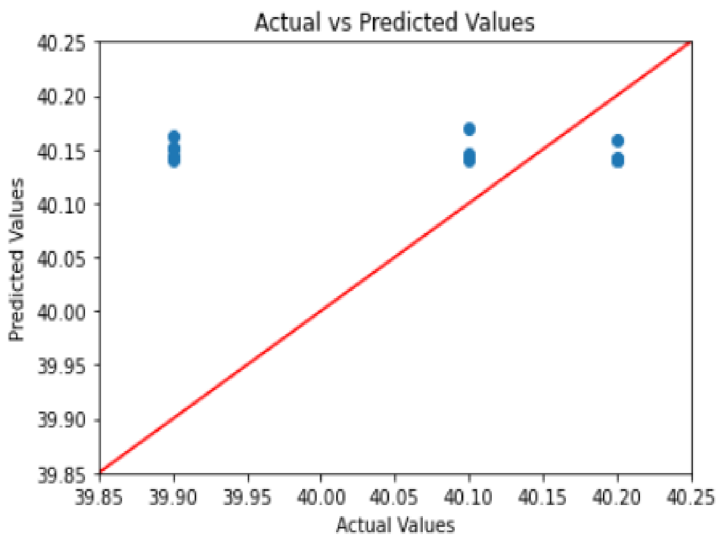
SAVI

LSTM RESULTS



MSE: 0.024816959516805227
RMSE: 0.1575339947973301

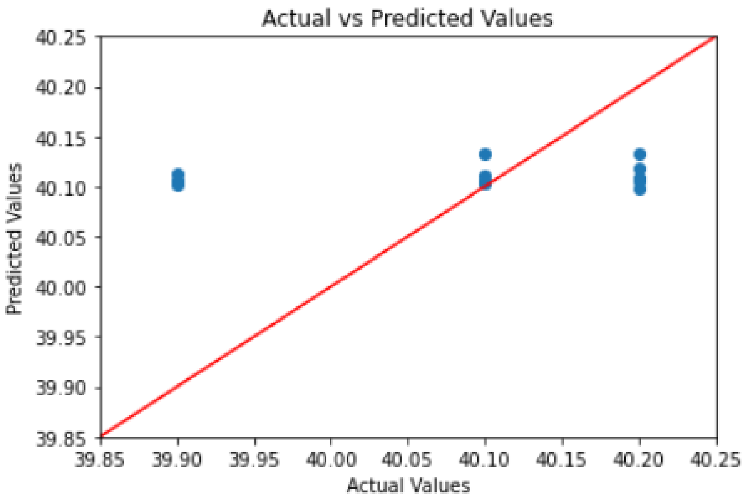
DVI



MSE: 0.022627755826300618
RMSE: 0.1504252499625665

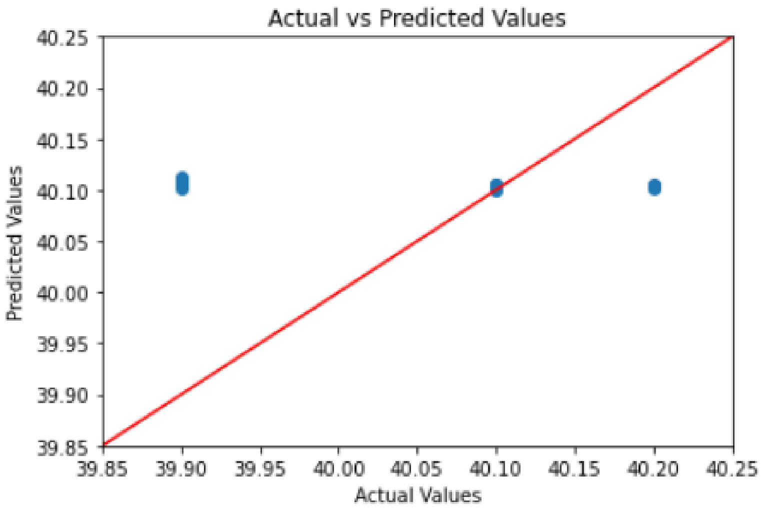
RVI

LSTM RESULTS



MSE: 0.016813471125982433
RMSE: 0.12966676955173378

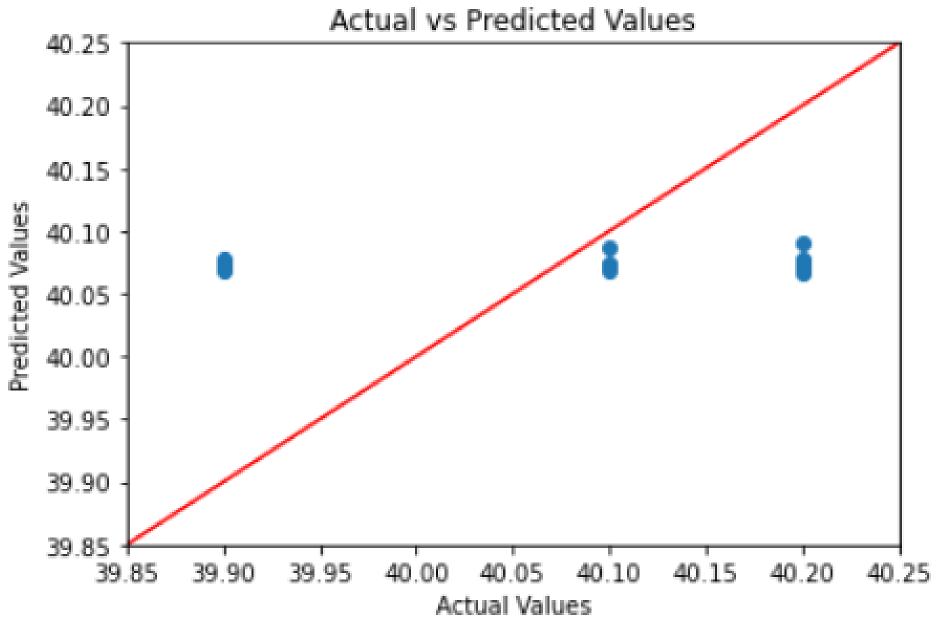
IPVI



MSE: 0.017382678931851005
RMSE: 0.1318433878958327

MSAVI

LSTM RESULTS




```
MSE:  0.015383224760881118
RMSE:  0.124029128679037
```

Lstm result Including all features

RESULTS

Features	Random forest (rmse)	lstm (rmse)
NDVI	0.13573865650716685	0.1324307245207583
SAVI	0.12979135500230213	0.13013934519347758
DVI	0.12998600752687917	0.1575339947973301
RVI	0.13399152949665274	0.1504252499625665
IPVI	0.16329389227595134	0.12966676955173378
MSAVI	0.12205492085001891	0.1318433878958327
ALL FEATURES	0.1280788593522111	0.124029128679037

CONCLUDING REMARKS

1.The LSTM model generally outperforms the random forest model in predicting the vegetation indices, as it has lower RMSE values for most of the features.

2.Among the individual features, MSAVI has the lowest RMSE values for both Random forest and LSTM models,indicating the most accurately predicted feature.

3.The combination of all features has the lowest RMSE value for the LSTM model,indicating that it is the best combination for predicting vegetation indices.

LIMITATIONS

- **Data availability :** One Of the primary limitations of the current Ml models for crop yield prediction is the lack of the data availability.Although several datasets are available for the crop yield prediction,they are often incomplete,inconsistent and not standardized.

- **Complexity of the model:** Many Ml models for crop yield prediction are complex and require a considerable amount of computational resources.this makes it difficult to deploy them in resource constrained environments such as rural areas.

- **Lack of Interpretability :** Many ML models used for crop yield prediction are black box models which means it is difficult to interpret the model's decision making process.This limits the ability of farmers to understand the factors that influence crop yields and the corrective measures.

FUTURE SCOPE

Integration of remote sensing and IoT data: The use of remote sensing and IoT data can provide more accurate and real time information about crop growth and other environmental factors that can influence crop yield. The integration these data sources with ML models can enhance their accuracy and reliability.

Developing Explainable AI models: The development of the explainable AI models that can provide insights into the decision-making process can help farmers make better decisions based on the model's recommendations collaborative efforts between researchers , farmers and industry can help improve the quality and availability of data for crop yield predictions. This can help develop more accurate and robust ML models for crop yield prediction.

these models should be easily accessible to farmers to improve crop yields and reduce input costs.

Contribution of Group Members

Group Members:

1.Murthati Mahibabu(20114058)

2.Priya(20114077)

3.Nimmagadda Vasavi (20114064)

Contributions:

Analytical approach - Vasavi and Priya

Simulation design - Mahibabu

Simulation modelling/algorithm - Mahibabu

Flow chart - Priya

Experimental setup - Vasavi and priya

Results/outcomes findings and validation - Vasavi

Report preparation - Vasavi and Priya

References

- **Kamir, E., Waldner, F. and Hochman, Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, pp.124-135.**
- **TY- JOUR, Mengjia,AU- He, Xiaohui,AU- Cheng, Xijie,AU-Li, Panle,AU- Luo, Haotian,AU-Tian, Zhihui,AU- Guo, Hengliang,PY - 2021/04/14 SP- 1,EP- 1,T1- Exploiting Hierarchical Features for Crop Yield Prediction Based on 3-D Convolutional Neural Networks and Multikernel Gaussian Process,VL - PP,DO - 10.1109/JSTARS.2021.3073149.**
- **E. Panwar, A. N. J. Kukunuri, D. Singh, A. K. Sharma and H. Kumar, "An Efficient Machine Learning Enabled Non-Destructive Technique for Remote Monitoring of Sugarcane Crop Health," in IEEE Access, vol. 10, pp. 75956-75970, 2022, doi: 10.1109/ACCESS.2022.3190716.**
- **Vasit Sagan, Maitiniyazi Maimaitijiang, Sourav Bhadra, Matthew Maimaitiyiming, Davis R. Brown, Paheding Sidike, Felix B. Fritschi,Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning,ISPRS Journal of Photogrammetry and Remote Sensing,Volume 174,2021,Pages 265-281,ISSN 0924-2716.**
- **Shirui Hao, Dongryeol Ryu, Andrew Western, Eileen Perry, Heye Bogena, Harrie Jan Hendricks Franssen,Performance of a wheat yield prediction model and factors influencing the performance: A review and meta-analysis,Agricultural Systems,Volume 194,2021,103278,ISSN 0308-521X.**
- **Mukesh Singh Boori, Komal Choudhary, Rustam Paringer, Alexander Kupriyanov, Machine learning for yield prediction in Fergana valley, Central Asia,Journal of the Saudi Society of Agricultural Sciences,Volume 22, Issue 2,2023,Pages 107-120,ISSN 1658-077X.**

References

- **Kundu, S.G., Ghosh, A., Kundu, A. and GP, G., 2022. A ML-AI ENABLED ENSEMBLE MODEL FOR PREDICTING AGRICULTURAL YIELD. *Cogent Food & Agriculture*, *8*(1), p.2085717.TY-CHAP,AU-Sinwar,Deepak,AU-Dhaka,Vijaypal,AU-Sharma,ManojKumar,AU-RaniGeeta,PY-2019/10/01,SP-155,EP-180,SN-978-981-15-0662-8,TI-AI-BasedYieldPredictionandSmartIrrigation,DO10.1007/978-981-15-0665_8.**
- **N. Gandhi, O. Petkar and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, 2016, pp. 105-110, doi: 10.1109/TIAR.2016.7801222.**
- **R. Luciani, G. Laneve and M. JahJah, "Agricultural Monitoring, an Automatic Procedure for Crop Mapping and Yield Estimation: The Great Rift Valley of Kenya Case," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2196-2208, July 2019, doi: 10.1109/JSTARS.2019.2921437.**
- **Martínez-Ferrer, L., Piles, M. and Camps-Valls, G., 2020. Crop yield estimation and interpretability with Gaussian processes. *IEEE Geoscience and Remote Sensing Letters*, *18*(12), pp.2043-2047. Cedric, L.S., Adoni, W.Y.H., Aworka, R., Zoueu, J.T., Mutombo, F.K., Krichen, M. and Kimpolo, C.L.M., 2022. Crops yield prediction based on machine learning models: case of west african countries. *Smart Agricultural Technology*, p.100049.**
- **Sathya P, Gnanasekaran P. (2023) Paddy Yield Prediction in Tamilnadu Delta Region Using MLR-LSTM Model. *Applied Artificial Intelligence* 37:1.**
- **Rafael Battisti, Paulo C. Sentelhas, Kenneth J. Boote,Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil,Field Crops Research, Volume 200,2017,Pages 28-37,ISSN 0378-4290.**

References

- **Bose, P., Kasabov, N.K., Bruzzone, L. and Hartono, R.N., 2016. Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(11), pp.6563-6573.**
- **Juan Cao, Zhao Zhang, Yuchuan Luo, Liangliang Zhang, Jing Zhang, Ziyue Li, Fulu Tao, Wheat yields predictions at a county and field scale with deep learning, machine learning, and google earth engine,European Journal of Agronomy,Volume 123,2021,126204,ISSN 1161-0301.**
- **Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L. and Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and electronics in agriculture*, *121*, pp.57-65.**
- **Diego Gómez, Pablo Salvador, Julia Sanz, José Luis Casanova,Modeling wheat yield with antecedent information, satellite and climate data using machine learning methods in Mexico, Agricultural and Forest Meteorology,Volume 300,2021,108317,ISSN 0168-1923.**
- **Rajkumar Dhakar, Vinay Kumar Sehgal, Debasish Chakraborty, Rabi Narayan Sahoo, Joydeep Mukherjee, Amor V.M. Ines, Soora Naresh Kumar, Paresh B. Shirsath, Somnath Baidya Roy, Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing,Agricultural Systems, Volume 195,2022,103299,ISSN 0308-521X.**
- **Glorot, X. and Bengio, Y., 2010, March. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.**