Evaluating Soybean Yield Using Multi - Model Ensembles in Osmanabad District (Kharif - 2023)

Priyanka Shamraj[1], Ashutosh Pawar[2], Upasana Singh[3], Bhargav Sonawane[4]

[1]Agrometeorologist, Department of GIS and Remote Sensing, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

[2] Head, GIS and RS, Department of GIS and Remote Sensing, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

[3]Sr. GIS and RS Executive, Department of GIS and Remote Sensing, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

[4]Jr. RS Executive, Department of GIS and Remote Sensing, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

Abstract: ***A study was conducted at the Semantic Technologies and Agritech Services, Pvt. Ltd., GIS, and Remote Sensing Team in Pune during the Kharif - 2023 season. The methodology outlined in the YESTECH manual, under the Pradhan Mantri Fasal Bima Yojana (PMFBY), was diligently followed. Osmanabad district has been experiencing significant weather - based yield losses in recent years. This case study aimed to estimate the yield of soybean crops for agricultural stakeholders, insurance companies, and government policies at the Revenue Circle level (RC). A multimodal approach was adopted over a single - model yield estimation approach to ensure the ensemble yield for precise forecasting/estimating of crop yield. The accuracy achieved was within a certain percentage, with Root Mean Square Error (RMSE) measured was less than ±30% at the RC level. Consequently, the overall findings suggest that employing such models for yield estimation represents one of the best approaches for decision - making among insurance stakeholders, particularly in rainfed regions where adverse impacts on soybean productivity have been observed under various climate change scenarios.***

Keywords: Remote Sensing, GIS, Net Primary Productivity (NPP), Machine Learning, Decision Support System for Agrotechnology Transfer (DSSAT - 4.8), Soybean, Osmanabad, Yield Simulation, Revenue Circle, Soybean Productivity.

1. Introduction

Agriculture stands as the cornerstone of global economies, supporting livelihoods and sustenance for billions worldwide. Accurately predicting crop yield is paramount for effective resource management, risk mitigation, and informed decision - making. Traditional methodologies, reliant on historical data and manual observations, often fail to address the dynamic nature of contemporary agricultural challenges. The integration of advanced technologies has ushered in a new era in agriculture, enabling a more nuanced and precise understanding of crop dynamics. Through the synergy of software applications, remote sensing, Geographic Information Systems (GIS), and Artificial Intelligence/Machine Learning (AI/ML) algorithms, vast datasets are processed, patterns analyzed, and crop yields predicted with unprecedented accuracy.

Unpredictable rainfall, rising temperatures, and extreme weather events such as hailstorms and strong winds pose threats to crop growth and yields. Farmers adapt by shifting to drought - resistant crops, increasing reliance on irrigation, combating soil degradation, and navigating economic vulnerability due to unstable production.

In the contemporary agricultural landscape, accurate estimation of crop yield has emerged as a critical aspect impacting various sectors including insurance, economy, government policies, and ultimately, the welfare of farmers. Traditional methods of crop yield estimation often suffer from limitations in accuracy and efficiency. However, the integration of advanced technologies such as sophisticated software, remote sensing, GIS, and cutting - edge AI/ML techniques has revolutionized the precision and reliability of crop yield estimation. Importance of Accurate Crop Yield Estimation:

1) Insurance Sector: Accurate crop yield estimates play a pivotal role in the insurance sector, facilitating precise risk assessment and the development of tailored insurance products.
2) Economic Implications: Crop yield estimates are fundamental to economic forecasting, influencing commodity markets, trade agreements, and pricing mechanisms.
3) Government Policies: Governments rely on accurate crop yield estimates to formulate effective agricultural policies, including the allocation of subsidies, resource distribution, and planning for strategic interventions during adverse weather conditions or pest outbreaks.
4) Public Welfare and Food Security: Accurate crop yield estimates are integral to ensuring food security and public welfare.
5) Farmers' Wellbeing: Precise crop yield estimates empower farmers to enhance planning and risk management. Access to reliable information enables informed decisions regarding crop selection, resource allocation, and market participation, thereby improving overall farm productivity and livelihoods.

The adoption of advanced methodologies for crop yield estimation represents a transformative step toward building agricultural resilience in the face of evolving challenges. The synergy between software applications, remote sensing, and GIS technologies empowers stakeholders across sectors to make informed decisions, fostering a sustainable and prosperous future for agriculture. By recognizing the multifaceted implications of accurate crop yield estimation, societies can collaboratively strengthen the foundations of global food security, economic stability, and the welfare of farming communities.

This research paper port delves into the significance of employing advanced methodologies for estimating crop yield and highlights their implications across diverse domains.

2. Material and Methods

Study area:
Study was carried out at Semantic Technologies and Agritech Services, Pvt. Ltd. Pune during *kharif* season 2023 for particular assignment. For this study, all revenue circles (RC) in the districts of Osmanabad of Maharashtra state were used as experimental sites. Field level data like ground truth, Crop cutting experiments were carried out.

Geography and Climate for Osmanabad District:
Osmanabad district covers an approximate area of 7, 546 square kilometres. It is located in the southeastern part of Maharashtra, near the Karnataka border. The district lies between latitudes 17°35′N to 18°40′N and longitudes 75°12′E to 76°25′E. Elevations within the district range from approximately 400 to 800 meters above sea level. Some part of the major rivers like Godawari and Bhima come under this district. The district receives an average annual rainfall of around 520 millimetres. During the Kharif season, which typically spans from June to September, Osmanabad receives between 350 - 390 millimetres of rainfall. Temperatures during the Kharif season range from 33 - 37°C maximum and 22 - 25°C minimum. The average relative humidity ranges from 70 - 80%. The predominant soil type in Osmanabad district is Vertisols. These soils are known for their high clay content and can pose drainage challenges due to their tendency to swell and shrink with moisture changes. Major crops cultivated in Osmanabad district include Soybean, Cotton, Jowar (Sorghum), Bajra (Pearl Millet), Tur (Pigeon Pea) and Sugarcane.

Methodology
All methodology was followed by the procedure given by yield estimation system based on technology (yes - tech) under Pradhan Mantri Fasal Bima Yojana (PMFBY).

Methodology used is multimodal approach for estimation of crop yield was given below. RC wise

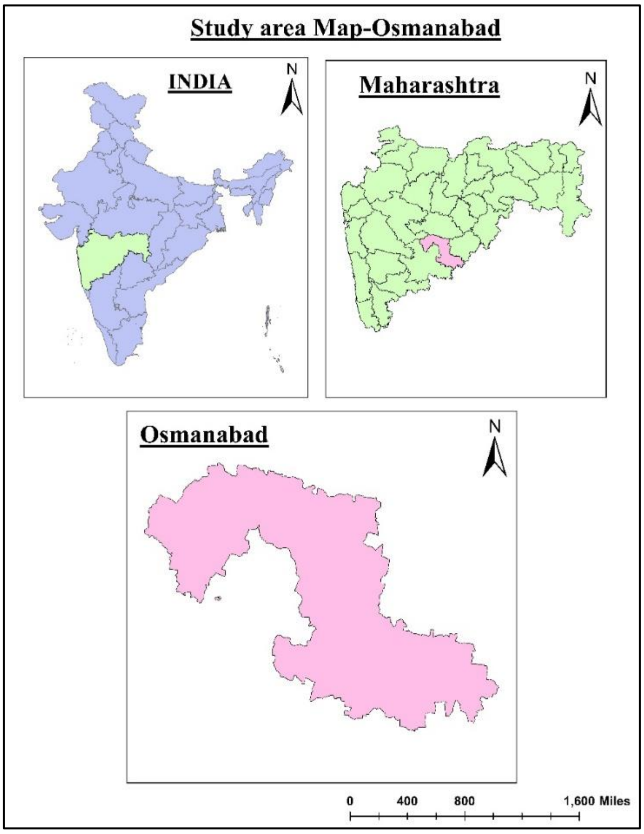

Figure 1: Study Area Map

yield in Tonnes/hector of soybean crop during *kharif* season 2023 was estimated by all following methods.

1) Semi Physical NPP - Net Primary Productivity
2) AI and Machine learning
3) Crop simulation model - DSSAT - 4.8
4) Ensemble Model

1) Semi Physical Net Primary Productivity (NPP):

Data and materials used:
The data and materials used in this study are as follows:

Table 1: Data used for NPP generation in Semi Physical model

Data	Satellite/Ground	Resolution	Source
Daily insolation/PAR	INSAT - 3D	4km resampled to 1km	MOSDAC
10 days composite fAPAR ver.2	PROBA V and SPOT - VGT	1km	Copernicus Land Service
8 days composite surface reflectance	Terra – MODIS	1km	MODIS Time Series Tool
Paddy Mask	Sentinel 1	5m	USGS Explorer
Temperature	Gridded data from NASA Power website	1km interpolated	NASA Power
Light - use efficiency			Literature
Harvest Index	Ground	CCE	

Methods: The different methods used in the study are as follows:

Fraction of Absorbed PAR (FAPAR):
The FAPAR data is from Copernicus Land Service, source link is (https: //land. copernicus. eu/global/index. html). the 10 - day composite product with 1 km data is used. The range of FAPAR lies between 0 and 1. The physical values are retrieved from the Digital Number (DN).

Photosynthetically Absorbed Radiation (PAR):
PAR is calculated from daily insolation data. The daily insolation data is converted to 8 - day composite (sum) for the whole period.50% insolation is considered as PAR. This daily insolation data is collected from MOSDAC from INSAT - 3D satellite, source link (www.mosdac. gov. in) for the crop season from 2018 to 2022.

*PAR= **8 - day composite * 0.5.***

Water Stress (Wstress):
The Wstress is calculated from Land Surface Water Index (LSWI). The MODIS time series tool (MODIStsp) used to download and process the MODIS 8 day composite (MOD09A1) source link is (https: //lpdaac. usgs. gov/products/mod09a1v006), and LSWI is calculated for the entire period with the formula

$LSWI = (\rho NIR - \rho SWIR) / (\rho NIR + \rho SWIR)$

LSWI value range from - 1 to 1, and higher positive values indicate the vegetation and soil water stress. Further, the Wstress is calculated from 8 days LSWI output –

$Wstess = (1 - LSWI) / (1 + LSWImax)$

The LSWImax value has been taken from the spatial maximum of particular crop mask of the entire district.

Temperature Stress:
Temperature Stress (Tstress): The daily average temperature data is downloaded from NASA Power website, source link is (https: //power. larc. nasa. gov/data - access - viewer. html). It is a gridded data with a resolution of 1°0 * 1°0 latitude and longitude.

$$T\ Stress = \frac{(T - Tmin) * (T - Tmax)}{[(T - Tmin) * (T - Tmax) - T - Topt)^{2}]}$$

Where, Tmin = Minimum temperature required for the photosynthesis (°C).
Tmax =Maximum temperature required for the photosynthesis (°C).
Topt = Optimal temperature required for the photosynthesis (°C);
T = Daily mean temperature (°C).

Table 2: Data used for soybean crop for Semi - Physical Approach.

Sr. No.	Particulars	Values	Source	Sr. No.	Particulars	Values	Source
1	T maximum	35°C	(Nimje, P. M.2022)	4	LUE	1.78	(Chavan et al., 2018)
2	T minimum	10°C		5	Harvest Index	0.45	Periodic CCE data.
3	T optimum	26°C					

On the off chance that air temperature falls beneath Tmin, which is quite a rare chance than Tscalar value will automatically become 0.

Light Use Efficiency (Ɛ):
The light use efficiency LUE is used for soybean crop was 1.78 for the study.

Crop Mask
The crop mask was derived utilizing Sentinel - 1 synthetic aperture radar (SAR) data obtained from the European Space Agency (ESA) Copernicus Hub. Employing the R programming language, we employed the Random Forest algorithm for the generation of the crop mask, implementing hyperparameter tuning techniques and contingency matrix analysis. This methodology was systematically applied across our specified crops within the targeted area of interest.

In terms of accuracy assessment, our results yielded a robust accuracy range of 90% to 95% across all cultivated crops and within various districts. This signifies a high level of precision in delineating and classifying the specified crops within the delineated geographical regions. The meticulous incorporation of Random Forest algorithm, hyperparameter tuning, and contingency matrix analysis has facilitated the generation of a reliable and accurate crop mask, providing valuable insights for agricultural monitoring and management within the designated study area.

Calculation of NPP and Grain Yield:
To compute the final Net Primary Productivity NPP and its Grain Yield, the formula and equation is used as follows. The NPP sum has been multiplied with Harvest Index (0.45) to estimate per pixel yield.

NPP = *PAR * FAPAR * Ɛ * Tstress * Wstress* (Logic of Monteith Equation 1972).

Same methodology is followed by Upasana Singh *et. al*. (2023) and also showing same results for all data used to run the model.

2) Crop simulation model - DSSAT
Material and method and all file process was carried out by the procedure followed by Hoogenboom, G., et. al (2019) and (2024) Jones, J. W., (2003) and the minimum data requirements for operation, calibration and validation of the Crop models are described below.

Crop simulation model is a mathematical equation or the set of equations, which represents the behaviour of system. We used CROPGRO – for Soybean crop. It is consisting of various subroutines *viz.*, Water balance subroutine, Phenology subroutine, Nitrogen subroutine, and Growth and Development subroutine described below.

Data input to model
The minimum data requirements for operation, calibration and validation of the Crop models are described below.

Table 3: Showing List of input required by crop simulation model

Sr. No.	Input variables	Acronym	Source
1.	**SITE DATA**		
	Latitude	LAT	NASA power
	Longitude	LONG	NASA power
	Elevation	ELEV	NASA power

Paper ID: SR24329123723 DOI: https://dx.doi.org/10.21275/SR24329123723

<table>
<tr><td>2.</td><td>DAILY WEATHER DATA</td><td></td><td></td></tr>
<tr><td></td><td>Maximum temperature</td><td>TEMPMAX</td><td>NASA power</td></tr>
<tr><td></td><td>Minimum temperature</td><td>TEMPMIN</td><td>NASA power</td></tr>
<tr><td></td><td>Solar radiation</td><td>SOLARAD</td><td>NASA power</td></tr>
<tr><td></td><td>Rainfall</td><td>RAIN</td><td>NASA power</td></tr>
<tr><td>3.</td><td>SOIL CHARACTERISTICS</td><td></td><td></td></tr>
<tr><td></td><td>Soil texture</td><td>SLTX</td><td rowspan="22">DSSAT website
Where Global gridded - soil profile dataset at 10 - km resolution was Developed for DSSAT - 4.8
Software crop simulation models.</td></tr>
<tr><td></td><td>Soil local classification</td><td>SLDESC</td></tr>
<tr><td></td><td>Soil depth</td><td>SLDP</td></tr>
<tr><td></td><td>Colour, moist</td><td>SCOM</td></tr>
<tr><td></td><td>Albedo (fraction)</td><td>SALB</td></tr>
<tr><td></td><td>Photosynthesis factor (0 to 1 scale)</td><td>SLPE</td></tr>
<tr><td></td><td>pH in buffer determination method</td><td>SMPX</td></tr>
<tr><td></td><td>Potassium determination method</td><td>SMKE</td></tr>
<tr><td></td><td>Horizon - wise</td><td></td></tr>
<tr><td></td><td>Lower limit drained</td><td>LL (L)</td></tr>
<tr><td></td><td>Upper limit drained</td><td>DUL (L)</td></tr>
<tr><td></td><td>Upper limit drained</td><td>SAT (L)</td></tr>
<tr><td></td><td>Saturated hydraulic conductivity</td><td>SWCN (L)</td></tr>
<tr><td></td><td>Bulk density moist</td><td>BD (L)</td></tr>
<tr><td></td><td>Organic carbon</td><td>OC (L)</td></tr>
<tr><td></td><td>Clay (<0.002 mm) `</td><td>CLAY (L)</td></tr>
<tr><td></td><td>Silt (0.05 to 0.002 mm)</td><td>SILT (L)</td></tr>
<tr><td></td><td>Coarse fraction (>2 mm)</td><td>STONES (L)</td></tr>
<tr><td></td><td>Total nitrogen</td><td>TOTN (L)</td></tr>
<tr><td></td><td>pH in buffer</td><td>PHKCL (L)</td></tr>
<tr><td></td><td>Cation exchange capacity</td><td>CEC (L)</td></tr>
<tr><td></td><td>Root growth factor 0 to 1</td><td>SHF (L)</td></tr>
<tr><td>4</td><td>MANAGEMENT DATA</td><td></td><td></td></tr>
<tr><td></td><td>Sowing date</td><td>YRPLT</td><td rowspan="17">Krishi - Dainandini Published by in Vasantrao Naik Marathwada Krishi Vidypeeth, Parbhani,</td></tr>
<tr><td></td><td>Plant population at seedling</td><td>PLNATS</td></tr>
<tr><td></td><td>Planting method (TP/direct seeded)</td><td>PLME</td></tr>
<tr><td></td><td>Row spacing</td><td>ROWSPS</td></tr>
<tr><td></td><td>Row direction (degree from north)</td><td>AZIR</td></tr>
<tr><td></td><td>Seed rate</td><td>SDWTRL</td></tr>
<tr><td></td><td>Sowing depth</td><td>SDEPTH</td></tr>
<tr><td></td><td>Irrigation dates</td><td>IDLAPL</td></tr>
<tr><td></td><td>Irrigation amount</td><td>AMT (J)</td></tr>
<tr><td></td><td>Method of irrigation</td><td>IRRCOD</td></tr>
<tr><td></td><td>Fertilizer application dates</td><td>FDAY (J)</td></tr>
<tr><td></td><td>Fertilizer amount N</td><td>ANFER</td></tr>
<tr><td></td><td>Fertilizer type</td><td>IFTYPE</td></tr>
<tr><td></td><td>Fertilizer application method</td><td>FERCOD</td></tr>
<tr><td></td><td>Fertilizer incorporation depth</td><td>DFERT</td></tr>
<tr><td></td><td>Tillage date</td><td>TDATE</td></tr>
<tr><td></td><td>Tillage implements</td><td>TIMPL</td></tr>
</table>

Input files

The files are organized into input, output and experiment performance data file. The experiment performance files are needed only when simulated results are to be compared with data recorded in a particular experiment. In some cases, they could be used as input files to reset some variable during the course of a simulation run. The input files are further divided into those dealing with the experiment, weather and soil and the characteristics of different genotypes. Similarly output files are also further divided into those dealing with the overview, summary, growth, water, carbon and nitrogen balance.

Soil properties directory file: The file SOIL. SOL contained the list of different soils with their physical and chemical properties.

Soil profile initial condition file: The soil profile initial condition file contained the initial values of soil water, soil reaction and soil nitrogen data pertaining to this situation was entered.

Irrigation management file: The Irrigation management file has the provision of date and amount per fixed irrigation (mm) applied depth (cm) of management. Irrigation data pertaining to this situation was entered.

Fertilizer management file: The fertilizer management file contained the date, form and amount of nitrogen application. Accordingly, information on fertilizer application was entered in the file.

Treatment management file: The treatment management file contained the description of each treatment under separate title and serial numbers. The file also contained dates of planting and emergence, plant population at seeding and at emergence, planting method, planting distribution, row spacing, row direction, planting depth, planting material, transplant age, plants per hill, dates of simulation beginning etc. All needed information was entered for all the treatments.

Crop cultivars directory file

For Soybean CRGRO048 contained the list of different cultivars with their genetic coefficients. The modified genetic coefficients viz., CSDVAR, PPSEN, EMG - FLW, FLW - FSD, FSD - PHM, WTPSD, SDPDVR, SDFDUR, PODDUR, THRESH, SDPRO and SDLIP is used. Variety selected was JS - 335 which is mostly used in this area.

The genetic coefficients are the most important parameters which represents the genetic characteristics of the cultivar and on which the crop phenology, biomass production partitioning and yield potential of the crop depends. However, the actual performance is controlled by the external factors also.

Running the crop model: Once, all the desired files were created carefully the model was run for all the crops cultivars. Each run of model created output files.

3) Machine learning:

Methodology and processing of model is described below in details.

Data Collection and Ground Truthing:

- Collect remote sensing data (optical and radar imagery) for the study area, covering the growing season of the crops.
- Ground truth data collection using field surveys using CropTech App (prepared by compony) for accurate calibration and validation.

Crop Mask Extraction:

- Pre - process the remote sensing data to correct for atmospheric interference and geometric distortions.
- Apply image enhancement techniques to improve the visual quality of the images.
- Employ supervised or unsupervised classification algorithms to extract crop masks for Soybean fields.

Generation of Spectral Indices and use of RADAR backscatter:

- Calculate vegetation indices (e. g., NDVI, NDRE, GNDVI) from the optical remote sensing data to assess crop health and Vigor.
- Utilize backscatter data from radar imagery to analyse surface roughness and other relevant crop information (VV, VH).

Crop Cutting Experiments:

- Use of Crop Cutting Experiment (CCE) for Crop with smart sampling methods to efficiently estimate crop parameters for crop.

Training and Testing Models (Machine Learning):

- Divide the dataset into training and testing sets, ensuring no overlap between the two.
- Evaluate the model's performance on the testing dataset using evaluation metrics like accuracy, F1 - score, and mean squared error (RMSE).

Model Validation and Final Result:

- Validate the trained model using independent ground truth data collected during the growing season for Soybean.
- Assess the model's accuracy and generalization ability to ensure reliable yield estimation.
- Obtain the final crop yield estimation results for Soybean in the study area.

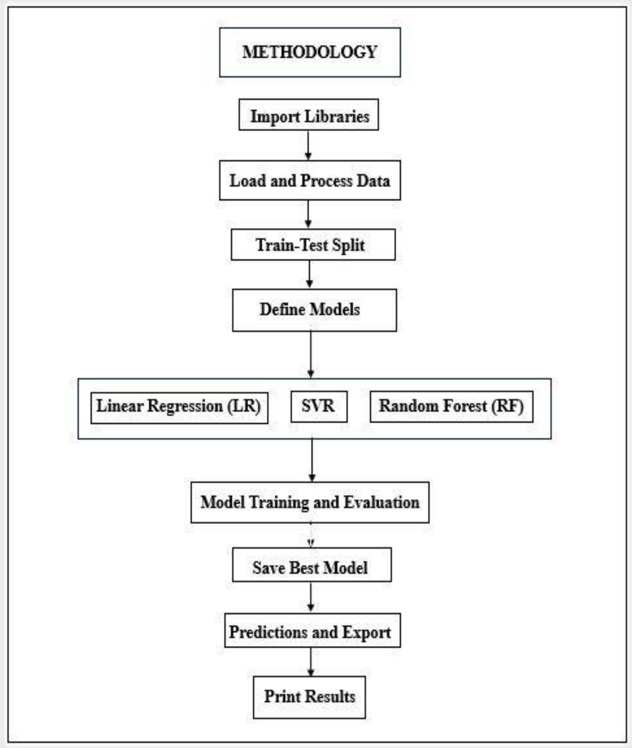

Figure 2: Methodology used in Machine learning Approach

4) Ensemble Models

This methodology aims to combine the predictive power of both Machine Learning (ML) models and Crop Simulation Models (CSM) to provide an enhanced and more accurate estimation of crop yields. Here is a structured approach:

1) Data Collection and Preprocessing:

- Gather data from both ML, Semi - Physical Approach and CSM approaches as outlined in the above methods.
- Consolidate all input data: weather data, soil properties, crop management practices, spectral indices, RADAR backscatter, and ground truth data.
- Ensure data alignment in terms of temporal and spatial granularity.

2) Individual Model Generation:

a) Machine Learning Approach:

- Utilize various algorithms like Linear regression, Random Forest, Extra Trees, k - earest neighbours, and neural networks.
- Train these models on the dataset ensuring proper validation and calibration.

b) Crop Simulation Approach:

- Use well - calibrated crop simulation models such as DSSAT.
- Simulate the growth and yield of crops using these models based on provided input data.

c) Semi - physical Models: A semi - physical model in remote sensing and GIS is a type of model that combines physical

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

principles with remotely sensed data to estimate or predict biophysical parameters, such as crop yield, biomass. These models are often used to monitor and manage natural resources, as well as to assess the impacts of climate change and other environmental stressors.

3) Ensemble Techniques Application:

- **Model Averaging**: Calculate the simple mean of predictions from ML, semi - physical model and CSM models.
- **Weighted Averaging**: Assign weights based on individual model performance and calculate the weighted average of predictions.
- **Stacking**: Use a meta - model that takes predictions from individual models as inputs and predicts the final yield.
- **Voting**: Each model votes for a final yield prediction, and the most frequent prediction is considered.

4) Model Validation:

- Split the dataset into training, validation, and test sets to avoid overfitting and ensure generalizability.
- Use metrics like Root Mean Squared Error (RMSE), and R - squared (R2) for evaluation.
- Assess performance using the test dataset and ground truth data.

5) Quality Control:

- Calculate the normalized RMSE between the observed and ensemble model's estimated yield.
- Ensure RMSE does not exceed acceptable thresholds, refining the model if necessary.

Validation:
The accuracy of our model was evaluated based on crop cutting experiment data (CCE data) of PMFBY (Pradhan Mantri Fasal Bima Yojana) for the crop season *kharif* - 2023.

3. Results and Discussion

Following were the results and conclusion for different methods/models used for estimation of yield of soyabean crop in Osmanabad districts of Maharashtra, Revenue - Circle wise.

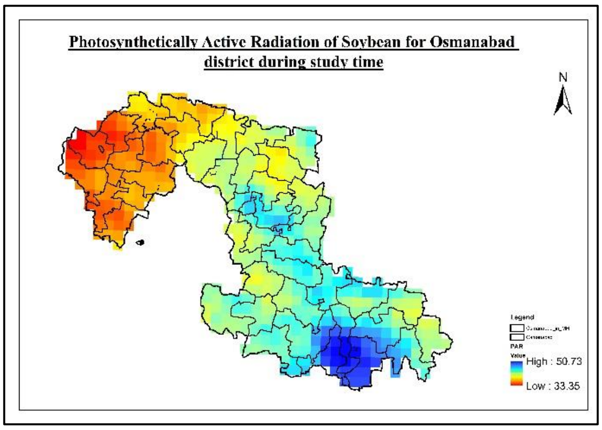

Figure 3: PAR for Osmanabad during *kharif* 2023

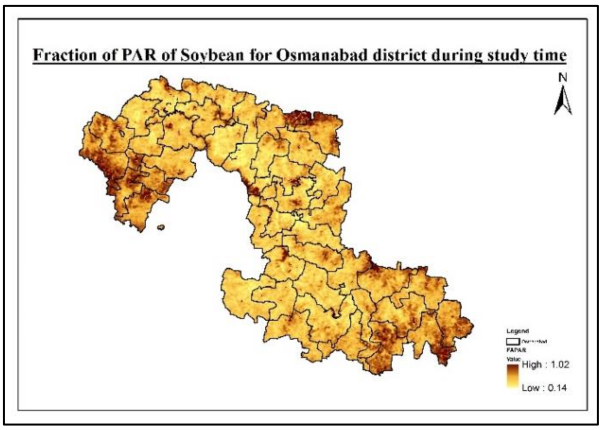

Figure 4: FAPAR for Osmanabad during *kharif* 2023

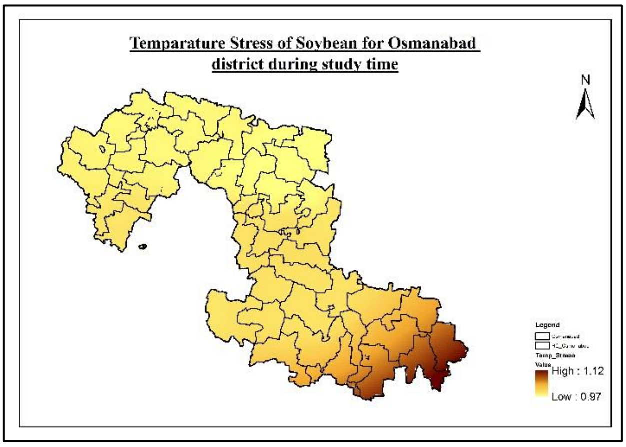

Figure 5: Tstress for Osmanabad during *kharif* 2023

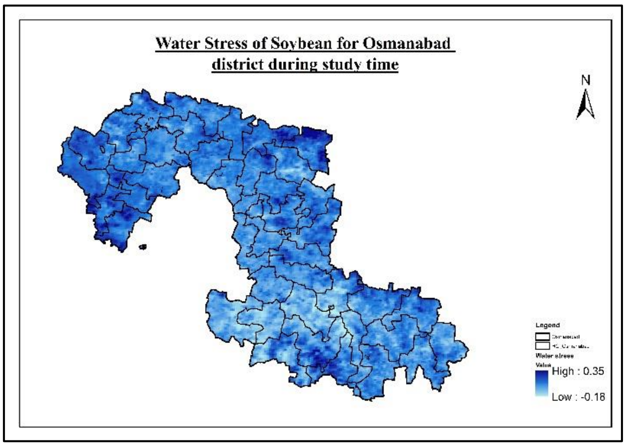

Figure 6: Waterstress for Osmanabad during *kharif* 2023

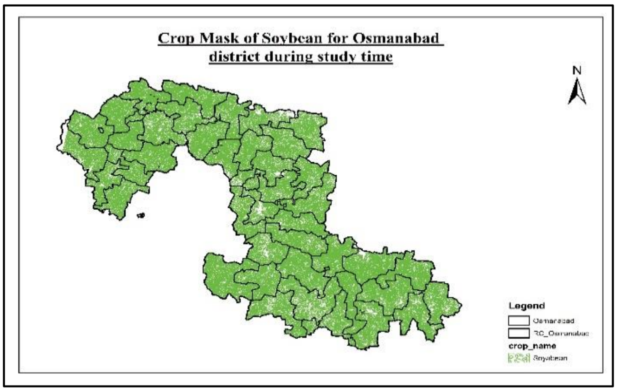

Figure 7: Soybean Crop Mask of Osmanabad during *kharif* 2023

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

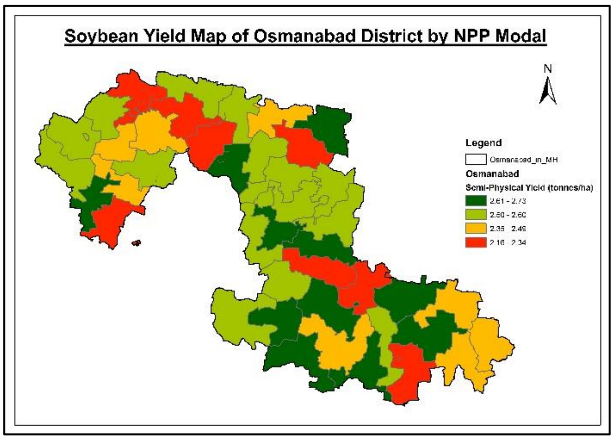

Figure 8: Soybean yield of Osmanabad during *kharif* 2023

1) Semi Physical Approach - NPP:

- The soybean crop yield in Osmanabad district for the year 2023 ranged from 0.71 to 1.70 tonnes per hectare.
- The average soybean crop yield in Osmanabad district for the year 2023 was 1.24 tonnes per hectare.
- 11 out of 36 Revenue circles have an actual yield greater than the average yield.
- Osmanabad City has the highest actual yield (1.70 tonnes per hectare) and Washi taluka has the lowest actual yield (0.71 tonnes per hectare) The semi - physical yield is consistently higher than the actual yield across all talukas. Same results were reported by Xiao, X., et. al (2006) and Yao, Y., et. al (2021)

2) Crop Simulation Model DSSAT - 4.8

- The DSSAT model projected yields ranging from 2.16 to 2.73 tonnes/ha across different regions.
- High performers in actual yields included Govindpur (1.70 tonnes/ha) and Jawala N. (1.62 tonnes/ha).
- Shiradhon demonstrated the highest DSSAT projected yield at 2.73 tonnes/ha.
- There's notable variability between actual yields and DSSAT projections, indicating potential discrepancies or modeling differences. Jadhav, S. D et. al (2018), Bhosale, A. D., et. al (2015) and Deshmukh, S. D., et. al (2013) also elaborated same results for soybean.

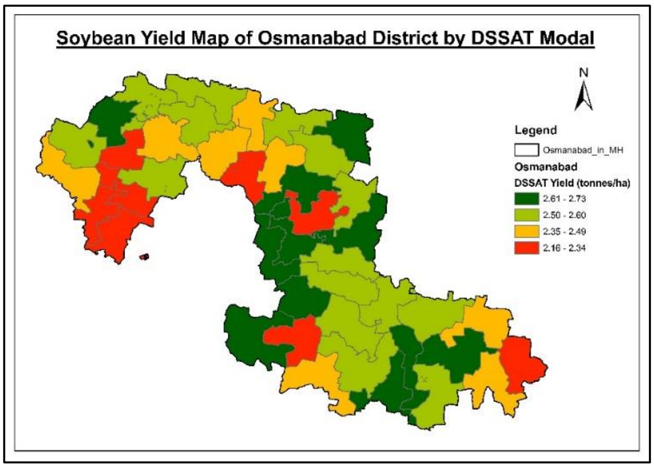

Figure 9: Soybean yield in T/ha by DSSAT for Osmanabad during *kharif* 2023

3) Machine learning

- CCE yield and different indices under study showing accuracy 82 % in Machine learning model. By the method (SVR) Support Vector Regression accuracy is showing highest value.
- The highest soybean crop yield in Osmanabad district for the year 2023 was in Ter RC, at 1.65 tonnes per hectare. While lowest in Washi RC, at 0.71 tonnes per hectare.
- 10 out of 34 revenue circles (RCs) have an average soybean crop yield higher than the district average.
- 24 out of 34 RCs have an average ML yield higher than the district average.
- ML yield refers to the marketable yield, which is the yield of the crop after it has been processed and is ready for sale.

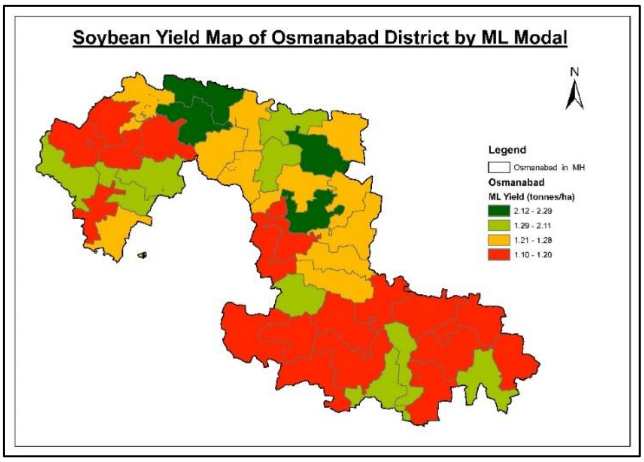

Figure 10: Soybean yield in T/ha by ML for Osmanabad during *kharif* 2023.

4) Ensemble Model:

- The Ensemble Yield represents a combination of all above three predictive models or methods to estimate soybean crop yield.
- Statistical approach give weightage during kharif 2023 as following to different models.

Model Used	DSSAT Yield	Semi - Physical Yield	Machine Learning Yield
Weightages in %	36.72	33.03	30.25

- Ensemble yields varied from 1.95 to 2.50 tonnes/ha across different regions.

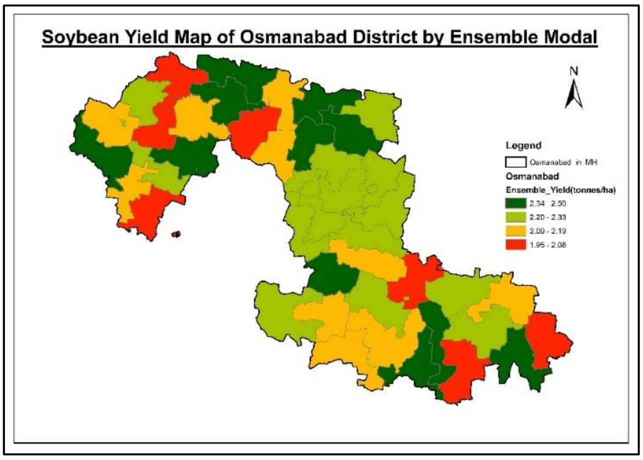

Figure 11: Soybean yield in T/ha by Ensemble Model for Osmanabad during *kharif* 2023.

- High performers in actual yields included Govindpur (1.70 tonnes/ha) and Jawala N. (1.62 tonnes/ha).
- Jewali demonstrated the highest ensemble yield at 2.50 tonnes/ha.
- There's variability between actual yields and ensemble yields, indicating differences in estimation methods or underlying factors affecting crop productivity. Same results were given by Md Didarul Islam et. al (2023), Liujun Xiao et. al. (2022) and Ayan Das a et. al (2023) in both Machine learning and ensemble approach.

Table 4: Estimated Yield of soybean crop in Tones/Hectors with different Models and percent error with ensemble model for year 2023.

Soybean crop yield in Osmanabad district for year 2023 in tonnes /ha								
District	Tehsil	RC	CCE	DSSAT Yield	Semi - Physical Yield	ML Yield	Ensemble Yield	RMSE% Error
Osmanabad	Paranda	Aasu	1.31	2.25	2.16	1.23	1.95	6
Osmanabad	Bhum	Ambhi	1.01	2.67	2.55	1.11	2.21	- 11
Osmanabad	Paranda	Anala	1.13	2.58	2.52	1.20	2.19	- 7
Osmanabad	Osmanabad	Bembli	1.17	2.55	2.67	1.26	2.25	- 8
Osmanabad	Bhum	Bhoom	0.95	2.49	2.49	1.19	2.14	- 26
Osmanabad	Umarga	Dalimb	1.39	2.66	2.66	1.14	2.26	18
Osmanabad	Osmanabad	Dhoki	1.35	2.62	2.51	1.25	2.22	7
Osmanabad	Bhum	Eit	1.17	2.51	2.25	1.21	2.07	- 4
Osmanabad	Kalamb	Govindpur	1.70	2.55	2.34	2.18	2.38	- 29
Osmanabad	Tuljapur	Itkal	1.45	2.48	2.65	1.16	2.19	20
Osmanabad	Kalamb	Itkur	1.38	2.36	2.51	1.28	2.13	7
Osmanabad	Osmanabad	Jagji	0.94	2.59	2.52	1.25	2.21	- 33
Osmanabad	Tuljapur	Jalkot	1.35	2.62	2.62	2.11	2.49	- 57
Osmanabad	Paranda	Jawala N.	1.62	2.23	2.49	2.07	2.28	- 28
Osmanabad	Lohara	Jewali	0.90	2.71	2.59	2.08	2.50	- 131
Osmanabad	Kalamb	Kalamb	1.54	2.51	2.48	2.07	2.38	- 34
Osmanabad	Osmanabad	Keshegaon	1.26	2.53	2.32	1.23	2.11	3
Osmanabad	Lohara	Lohara	1.42	2.55	2.23	1.19	2.07	16
Osmanabad	Lohara	Makni	1.22	2.52	2.62	1.20	2.20	2
Osmanabad	Tuljapur	Mangrul	1.03	2.32	2.68	1.13	2.13	- 10
Osmanabad	Bhum	Mankeshwar	1.58	2.52	2.58	2.07	2.42	- 31
Osmanabad	Kalamb	Moha	1.21	2.48	2.55	2.09	2.40	- 72
Osmanabad	Umarga	Mulaj	1.25	2.16	2.39	1.17	1.98	6
Osmanabad	Umarga	Murum	1.43	2.52	2.27	1.20	2.08	16
Osmanabad	Tuljapur	Naldurag	1.14	2.56	2.36	1.17	2.12	- 3
Osmanabad	Umarga	Narangwadi	1.05	2.49	2.46	1.15	2.12	- 10
Osmanabad	Osmanabad	Osmanabad C	1.10	2.67	2.71	1.20	2.29	- 9
Osmanabad	Osmanabad	Osmanabad R	1.41	2.72	2.53	1.19	2.25	16
Osmanabad	Osmanabad	Padoli	1.03	2.65	2.55	1.25	2.24	- 20
Osmanabad	Paranda	Paranda	1.20	2.34	2.67	1.16	2.14	3
Osmanabad	Washi	Pargaon	0.84	2.60	2.55	2.24	2.49	- 167
Osmanabad	Tuljapur	Salgara D.	1.19	2.51	2.73	1.15	2.23	3
Osmanabad	Tuljapur	Savargaon	1.22	2.68	2.60	1.14	2.24	7
Osmanabad	Kalamb	Shiradhon	1.39	2.73	2.66	1.24	2.31	11
Osmanabad	Paranda	Sonari	0.80	2.39	2.56	2.08	2.37	- 160
Osmanabad	Osmanabad	Ter	1.65	2.27	2.52	2.18	2.33	- 32
Osmanabad	Washi	Terkheda	1.21	2.46	2.16	1.27	2.04	- 5
Osmanabad	Tuljapur	Tuljapur	0.86	2.66	2.51	2.01	2.44	- 135
Osmanabad	Umarga	Umarga	1.48	2.45	2.45	2.07	2.35	- 40
Osmanabad	Bhum	Walwad	1.20	2.16	2.48	1.10	1.99	8
Osmanabad	Washi	Washi	0.71	2.55	2.24	2.29	2.37	- 223
Osmanabad	Kalamb	Yermala	1.23	2.24	2.72	1.27	2.15	- 3
Average yield =			**1.23**	**2.50**	**2.50**	**1.48**	**2.23**	**- 27**

In Table 4, the yield estimated by various methods is presented, including the percentage error of yield by the Machine learning model with field CCE, which is provided in the last column. Out of 42 points only 12 points were showing more than ±30% error. As per mentioned in deliverables in YESTECH manual given by Pradhan Mantri Fasal Bima Yojana, the error (nRMSE) between the observed and modeled yield should not be more than ±30% for district level. Which indicates that the process adopted for RC wise yield estimation is acceptable for all the models in Osmanabad district.

4. Conclusion

"The comparative evaluation of NPP, DSSAT, and Machine Learning models for predicting soybean crop yields in Osmanabad, Maharashtra for kharif 2023 has provided a nuanced understanding of their individual strengths and limitations. Each model demonstrated distinct capabilities in capturing the complexities of crop growth dynamics, with Machine Learning showcasing its adaptability and predictive accuracy.

The ensemble model, combining NPP, DSSAT, and Machine Learning, offered a holistic perspective by leveraging the strengths of individual models. This integration allowed for a more robust and reliable prediction of crop yields, providing a comprehensive overview of the crop performance over the study period.

Comparisons between the ensemble model results and field data revealed a promising alignment, emphasizing the potential of ensemble modelling in enhancing the accuracy of yield predictions. The combined approach contributes to minimizing uncertainties associated with individual models and provides a more reliable basis for decision - making in agriculture.

In conclusion, the integration of NPP, DSSAT, and Machine Learning models into an ensemble framework presents a promising avenue for advancing crop yield prediction methodologies. This study serves as a foundation for further research and refinement, with the ultimate goal of providing farmers and policymakers with accurate and actionable insights for sustainable agricultural practices in Maharashtra. "

References

[1] Ayan Das, Mukesh Kumar, Amit Kushwaha, Rucha Dave, Kailash Kamaji Dakhore, Karshan Chaudhari, & Bimal Kumar Bhattacharya. (2023). Machine learning model ensemble for predicting sugarcane yield through synergy of optical and SAR remote sensing. Remote Sensing Applications: Society and Environment, Volume 30, April 2023, 100962. https: //doi. org/10.1016/j. rsase.2023.100962

[2] Bhosale, A. D., Waskar, D. P., & Shinde, P. B. (2015). Performance of DSSAT model for simulating soybean yield under rainfed condition in Vertisols of central Maharashtra. International Journal of Agriculture, Environment and Biotechnology, 8 (3), 604 - 610.

[3] Chavan, K. K., Khobragade, A. M., Kadam, Y. E. and Mane, R. B. (2018) Study the heat unit requirement of soybean (Glycine max) varieties under varied weather condition at Parbhani. Journal of Pharmacognosy and Phytochemistry.7 (3): 526 - 530.

[4] Deshmukh, S. D., Waskar, D. P., & Shinde, P. B. (2013). Application of DSSAT model for soybean yield prediction in Vertisols of western Maharashtra. International Journal of Current Microbiology and Applied Sciences, 2 (8), 555 - 562.

[5] Hoogenboom, G., C. H. Porter, K. J. Boote, V. Shelia, P. W. Wilkens, U. Singh, J. W. White, S. Asseng, J. I. Lizaso, L. P. Moreno, W. Pavan, R. Ogoshi, L. A. Hunt, G. Y. Tsuji, and J. W. Jones. (2019). The DSSAT crop modeling ecosystem. In: p.173 - 216 [K. J. Boote, editor] Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom. https: //dx. doi. org/10.19103/AS.2019.0061.10

[6] Hoogenboom, G., C. H. Porter, V. Shelia, K. J. Boote, U. Singh, W. Pavan, F. A. A. Oliveira, L. P. Moreno - Cadena, T. B. Ferreira, J. W. White, J. I. Lizaso, D. N. L. Pequeno, B. A. Kimball, P. D. Alderman, K. R. Thorp, S. V. Cuadra, M. S. Vianna, F. J. Villalobos, W. D. Batchelor, S. Asseng, M. R. Jones, A. Hopf, H. B. Dias, L. A. Hunt, and J. W. Jones. (2024). Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.8.2 (www.DSSAT. net). DSSAT Foundation, Gainesville, Florida, USA.

[7] Jadhav, S. D., Waskar, D. P., & Shinde, P. B. (2018). Evaluation of DSSAT model for soybean yield prediction under different sowing dates and irrigation levels in Vidarbha region of Maharashtra. Journal of AgriSearch, 6 (4), 37 - 42.

[8] Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18, 235 - 265. https: //doi. org/10.1016/S1161 - 0301 (02) 00107 - 7

[9] Liujun Xiao, Guocheng Wang, Hangxin Zhou, Xiao Jin, & Zhongkui Luo. (2022). Coupling agricultural system models with machine learning to facilitate regional predictions of management practices and crop production. Environ. Res. Lett., 17 (2022), 114027. https: //doi. org/10.1088/1748 - 9326/ac9c71

[10] Md Didarul Islam, Liping Di, Faisal Mueen Qamer, Sravan Shrestha, Liying Guo, Li Lin, Timothy J. Mayer, & Aparna R. Phalke. (2023). Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning. Remote Sensing, 15 (9), 2374. Remote Sensing | Free Full - Text | Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning (mdpi. com)

[11] Monteith, J. L. (1982). Solar radiation and productivity in tropical ecosystems. Journal of Applied Ecology, 19 (3), 657 - 666.

[12] Nimje, P. M. (2022) Soybean Production technology, National Skill Development Corporation AISECT, Agriculture Skill Council of India 2022. (https: //www.nsdcindia. org/scmp/assets/image/1913597640 - 20_SoybeanProductionTechnology_preview. pdf)

[13] Upasana Singh, Gargi Gaydhane and Ashutosh Pawar (2023) A Semi - Physical Approach using Remote Sensing based Net Primary Productivity (NPP), Spatial, Spectral & Temporal Paddy Yield Model Development for the State of Assam. International Journal of Science and Research (IJSR).12 (8), 1175 - 1785.

[14] VNMKV Diary 2023 Krishi Dainandini 2023 published by Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani during 2023.

[15] Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J. Y., Salas, W., & Moore III, B. (2006). Mapping paddy rice agriculture in South and Southeast Asia using multi - temporal MODIS images. Remote Sensing of Environment, 100 (1), 95 - 113.

[16] Yao, Y., Li, Z., Tian, F., & Tao, F. (2021). Remote Sensing - Based Estimation of Maize Yield Using a Semi - Physical Approach: A Case Study in the North China Plain. Frontiers in Plant Science, 12, 662.