Volume 9, Issue 3, March – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

Multi-Model Ensemble Approach for Soybean Crop Yield Estimation (*Kharif*-2023) in Latur District at Macroscale level

Ashutosh Pawar
Head, GIS and RS, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

Upasana Singh
Sr. GIS and RS Executive, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

Priyanka Shamraj
Agrometeorologist, Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

Bhargav Sonawane
Jr. RS Executive Semantic Technologies and Agritech Services Pvt. Ltd., Pune.

Abstract:- Crop area estimation is a critical aspect of agricultural monitoring and management, providing essential information for decision-making in the agricultural sector. Study was carried out at Semantic Technologies and Agritech services Pvt. Ltd., GIS and Remote sensing team, Pune during Kharif-2023. All methodology given by YESTECH manual under Pradhan Mantri Fasal Bima Yojana (PFMBY) was followed. Latur district facing more weather-based yield losses during last few of years. In this case study we tried to estimate yield of soybean crop for agriculture-based stake holders, insurance companies, Government polices at Revenue circle level (RC). Multimodal approach is beneficial over single model yield estimation approach as it takes ensemble yield for perfect forecasting of crop yield. Accuracy was in the range as given in YESTECH manual at RC level. Thus, overall results show that use of such model for yield estimation is one of the best approach to take the decisions for insurance based stake holders in rainfed regions where more negative consequences on soybean productivity under different climate change scenario was observed.

Keywords:- *Remote Sensing, GIS, NPP, Machine Learning, DSSAT-4.8, Soybean, Latur, Yield Simulation, Revenue Circle, Soybean productivity.*

I. INTRODUCTION

Agriculture is the backbone of global economies, providing sustenance and livelihoods for billions of people. The ability to accurately predict crop yield is paramount for effective resource management, risk mitigation, and informed decision-making. Traditional methods, reliant on historical data and manual observations, often fall short in addressing the dynamic nature of modern 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. Software applications, remote sensing, GIS, and AI/ML algorithms work synergistically to process vast datasets, analyse patterns, and predict crop yields with unprecedented accuracy.

Unpredictable rainfall, rising temperatures, and extreme weather events like hailstorms and strong winds threaten crop growth and yields. Farmers shift to drought-resistant crops, rely heavily on irrigation, face soil degradation, and experience economic vulnerability due to unstable production.

In the contemporary agricultural landscape, the accurate estimation of crop yield has emerged as a critical aspect influencing various sectors including insurance, economy, government policies, and ultimately, the welfare of farmers. Traditional methods of crop yield estimation are often marred by limitations in accuracy and efficiency. However, the integration of advanced technologies, such as sophisticated software, remote sensing, GIS (Geographic Information System), and cutting-edge artificial intelligence (AI) and machine learning (ML) techniques, has revolutionized the precision and reliability of crop yield estimation.

➢ *Importance of Accurate Crop Yield Estimation:*

- Insurance Sector: Accurate crop yield estimates play a pivotal role in the insurance sector, enabling precise risk assessment and facilitating the development of tailored insurance products.
- Economic Implications: Crop yield estimates are fundamental to economic forecasting, impacting commodity markets, trade agreements, and pricing mechanisms.
- Government Policies: Governments rely on accurate crop yield estimates to formulate effective agricultural policies. This includes allocation of subsidies, distribution of resources, and planning for strategic interventions during periods of adverse weather conditions or pest outbreaks.
- Public Welfare and Food Security: Accurate crop yield estimates are integral to ensuring food security and public welfare.
- Farmers' Wellbeing: For farmers, precise crop yield estimates translate into enhanced planning and risk management. Access to reliable information empowers farmers to make informed decisions regarding crop

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

selection, resource allocation, and market participation, thereby improving overall farm productivity and livelihoods.

The adoption of advanced methods for crop yield estimation is a transformative step towards 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 work collaboratively to strengthen the foundations of global food security, economic stability, and the welfare of farming communities.

This report delves into the significance of employing advanced methods for estimating crop yield and highlights their implications across diverse domains.

II. MATERIAL AND METHODS

A. 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 Latur of Maharashtra state were used as experimental sites. Field level data like ground truth, Crop cutting experiments were carried out.

B. Geography and Climate for Latur District:

Latur districts cover 7,157 sq km of area. Annual rainfall averages 520 mm, with Kharif season receiving 350-390 mm. Kharif temperatures range from 33-37°C maximum and 22-25°C minimum, with average relative humidity of 70-80% located at latitudes 17°52′N to 18°50′N, 75°16′E to 76°42′E. Elevations range from 400-800 m. Soil and Drainage can be described as , Vertisols dominate the region, posing drainage challenges due to flat topography and leading to waterlogging during heavy rainfall. Situated in Boundary of Karnataka-Maharashtra border, the district is surrounded by Osmanabad in the South, Beed in the West, Parbhani and Nanded in the North and Bidar district of Karnataka in the East. Soybean, cotton, jowar, bajra, tur, and sugarcane is one of the major crops which is taken in this district.

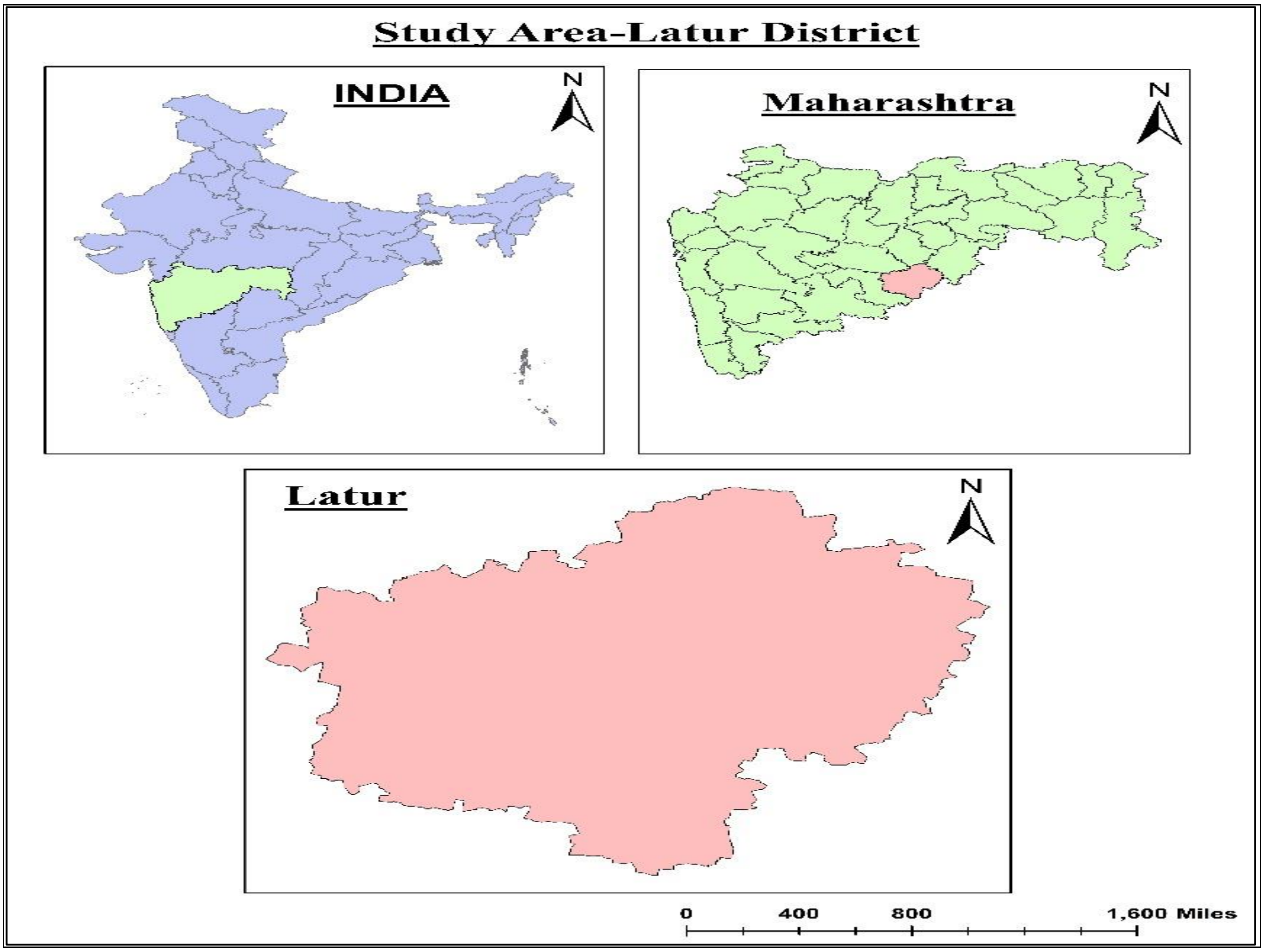

Fig 1: Study Area

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

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

- Semi Physical NPP- Net Primary Productivity
- AI and Machine learning
- Crop simulation model-DSSAT-4.8
- Ensemble Model

➢ *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	

- *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					

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

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.

➢ *Crop Simulation Model-DSSAT*

Crop simulation model is a mathematical equation or the set of equations, which represents the behavior 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*

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.

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
2.	**DAILY WEATHER DATA**		
	Maximum temperature	TEMPMAX	NASA power
	Minimum temperature	TEMPMIN	NASA power
	Solar radiation	SOLARAD	NASA power
	Rainfall	RAIN	NASA power
3.	**SOIL CHARACTERISTICS**		
	Soil texture	SLTX	DSSAT website Where Global gridded-soil profile dataset at 10-km resolution was Developed for DSSAT-4.8 Software crop simulation models.
	Soil local classification	SLDESC	
	Soil depth	SLDP	
	Colour, moist	SCOM	
	Albedo (fraction)	SALB	
	Photosynthesis factor (0 to 1 scale)	SLPE	
	pH in buffer determination method	SMPX	
	Potassium determination method	SMKE	
	Horizon-wise		
	Lower limit drained	LL(L)	
	Upper limit drained	DUL(L)	
	Upper limit drained	SAT(L)	
	Saturated hydraulic conductivity	SWCN(L)	
	Bulk density moist	BD(L)	
	Organic carbon	OC(L)	
	Clay (<0.002 mm) `	CLAY(L)	
	Silt (0.05 to 0.002 mm)	SILT(L)	

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

	Coarse fraction (>2 mm)	STONES(L)	
	Total nitrogen	TOTN(L)	
	pH in buffer	PHKCL(L)	
	Cation exchange capacity	CEC(L)	
	Root growth factor 0 to 1	SHF(L)	
4	**MANAGEMENT DATA**		
	Sowing date	YRPLT	Krishi-Dainandini Published by in Vasantrao Naik Marathwada Krishi Vidypeeth, Parbhani,
	Plant population at seedling	PLNATS	
	Planting method (TP/direct seeded)	PLME	
	Row spacing	ROWSPS	
	Row direction (degree from north)	AZIR	
	Seed rate	SDWTRL	
	Sowing depth	SDEPTH	
	Irrigation dates	IDLAPL	
	Irrigation amount	AMT(J)	
	Method of irrigation	IRRCOD	
	Fertilizer application dates	FDAY(J)	
	Fertilizer amount N	ANFER	
	Fertilizer type	IFTYPE	
	Fertilizer application method	FERCOD	
	Fertilizer incorporation depth	DFERT	
	Tillage date	TDATE	
	Tillage implements	TIMPL	

- *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.

- *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.

- ➢ *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.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

- *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.

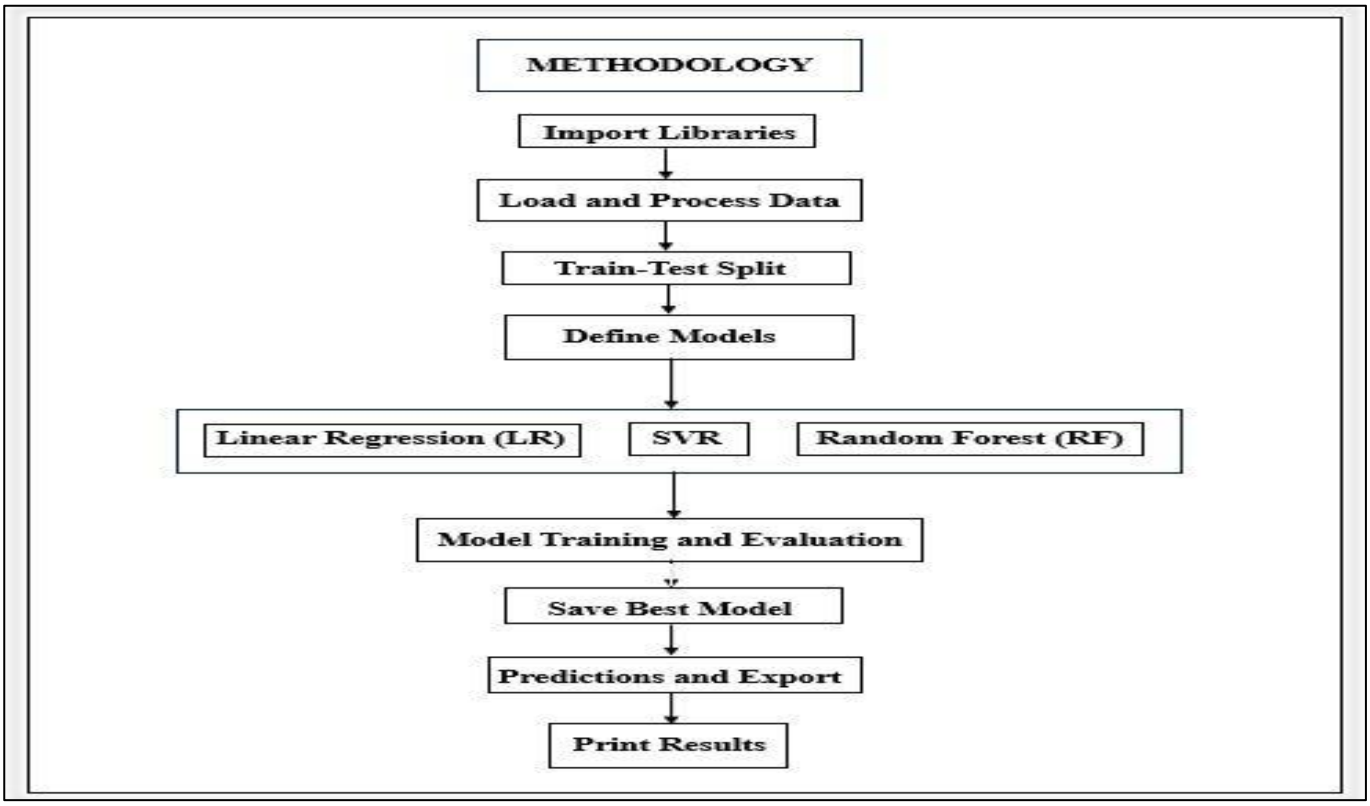

Fig 2: Methodology used in Machine learning Approach

➢ *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:

- *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.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

- *Individual Model Generation:*

✓ 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.

✓ 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.

✓ Semi-physical Models:

A semi-physical model in remote sensing and GIS is a type of model that combines physical 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.

- *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.

- *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.

- *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.

III. RESULTS AND DISCUSSION

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

A. Semi Physical Approach-NPP:

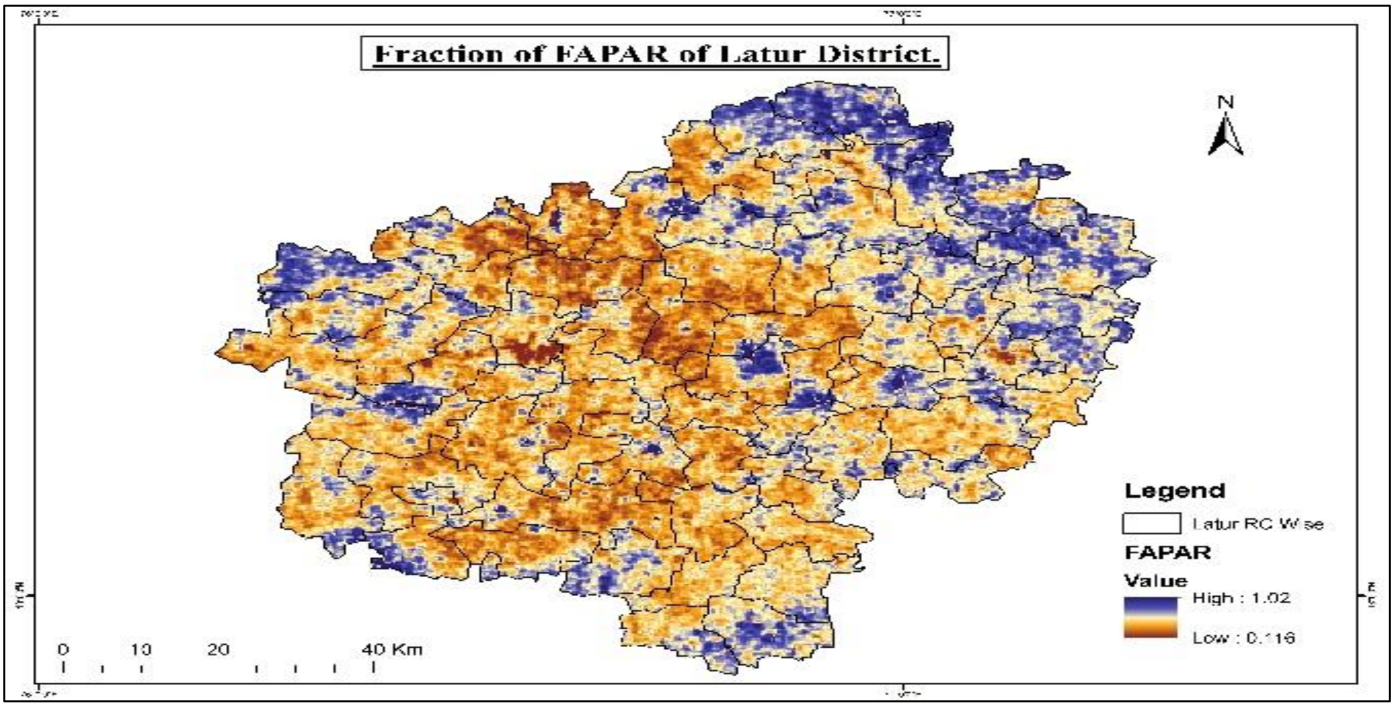

Fig 3: FAPAR for Latur during *Kharif* 2023

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

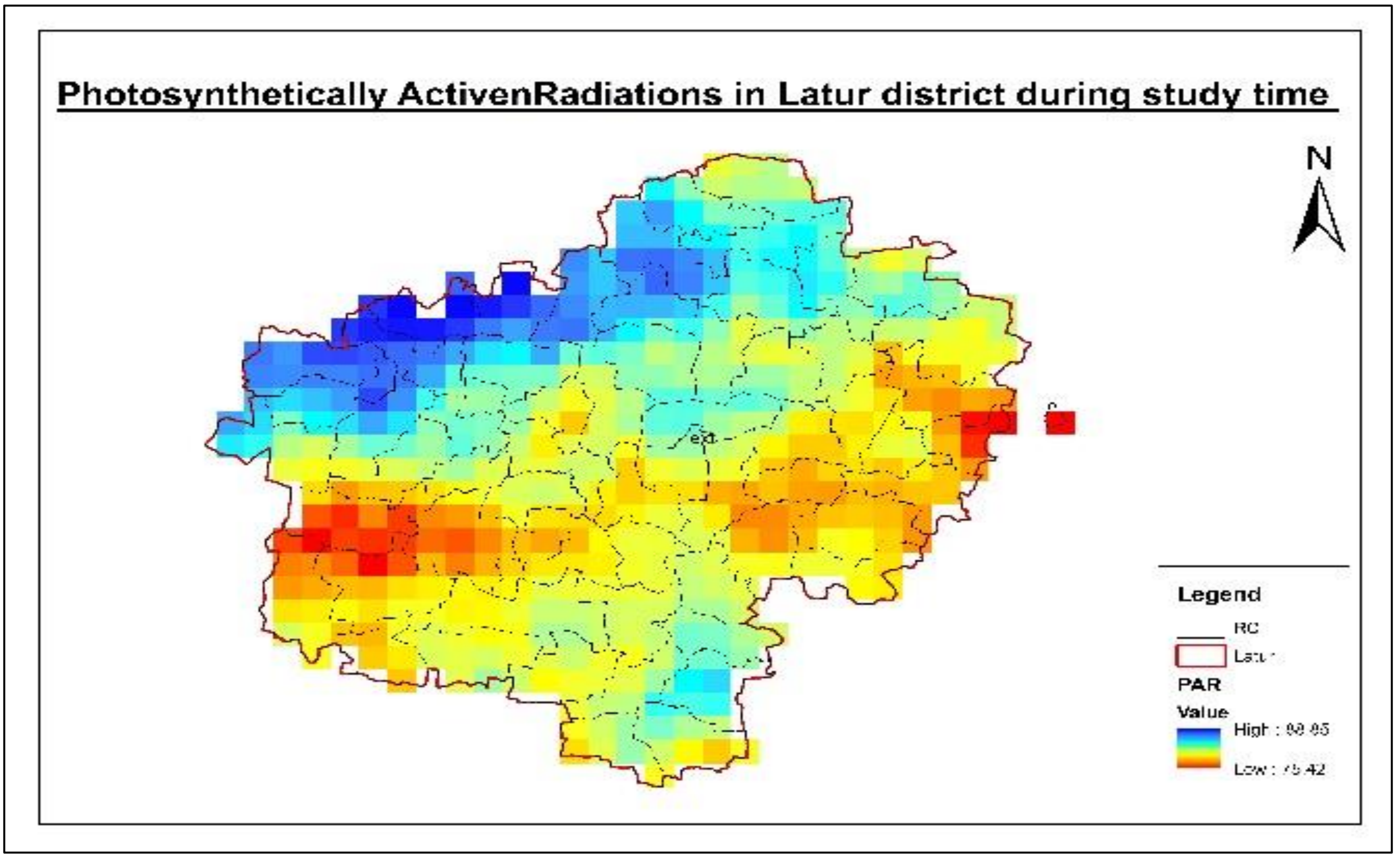

Fig 4: PAR for Latur during *Kharif* 2023

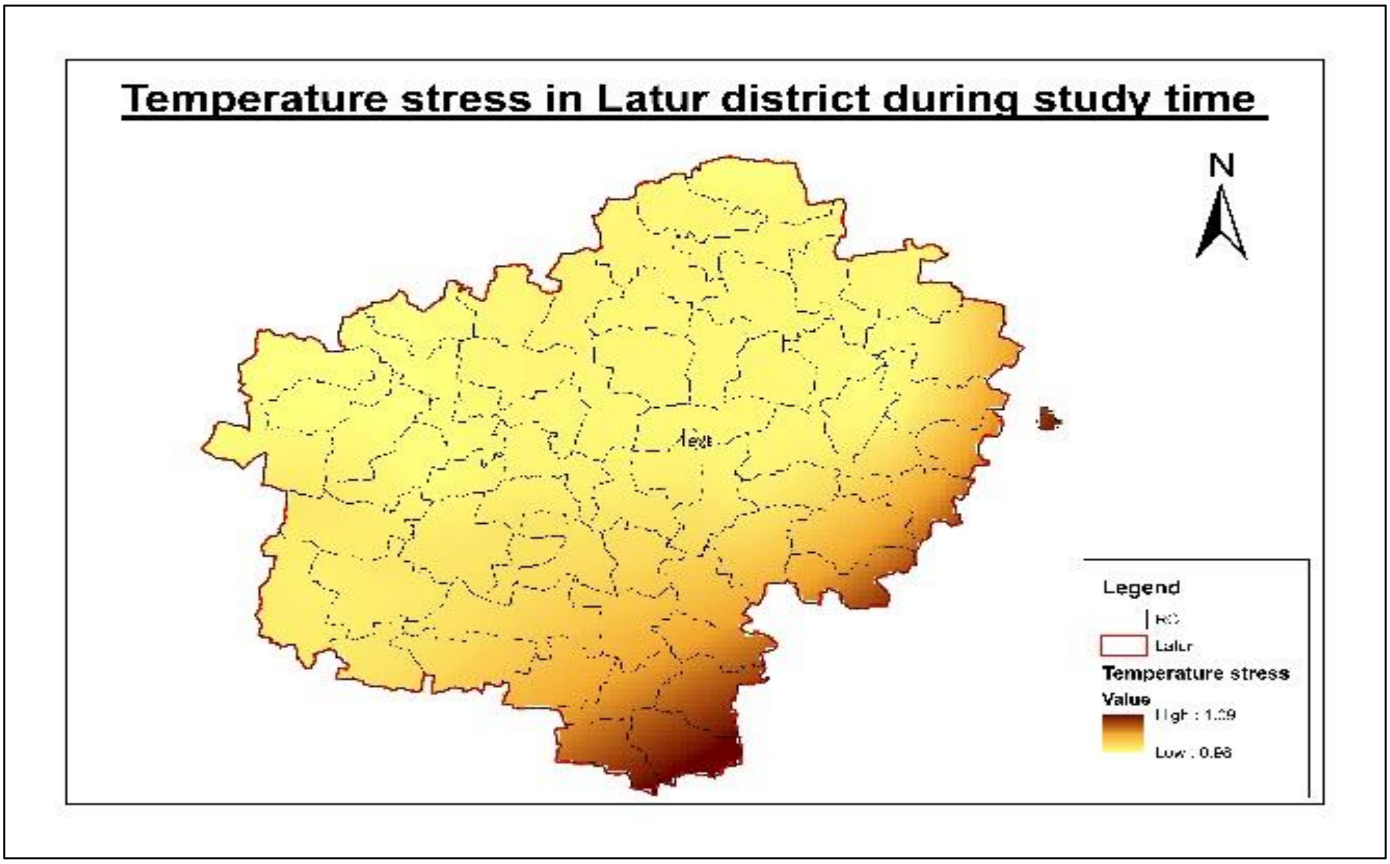

Fig 5: Tstress for Latur during Kharif 2023

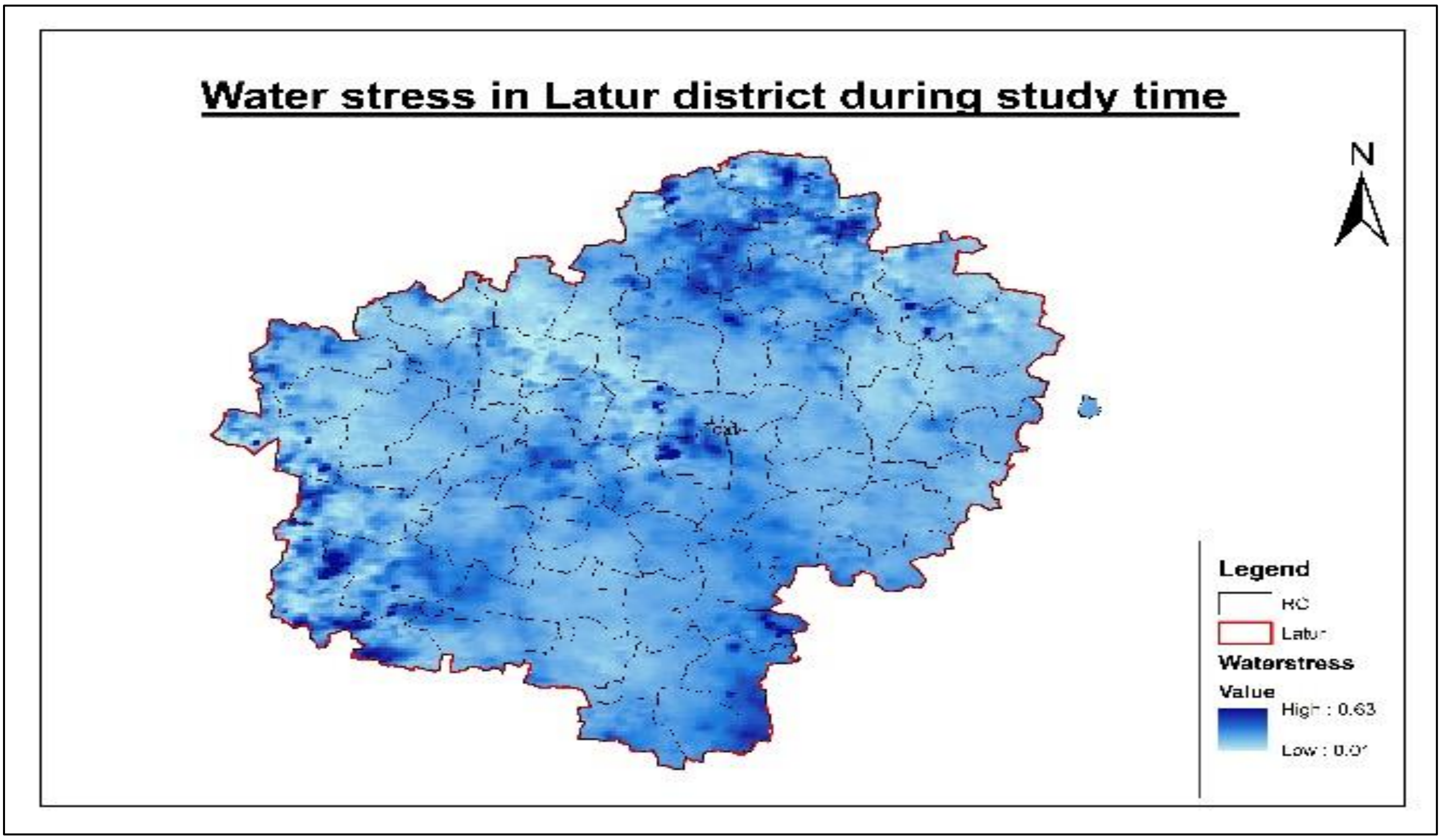

Fig. 6: Waterstress for Latur during Kharif 2023

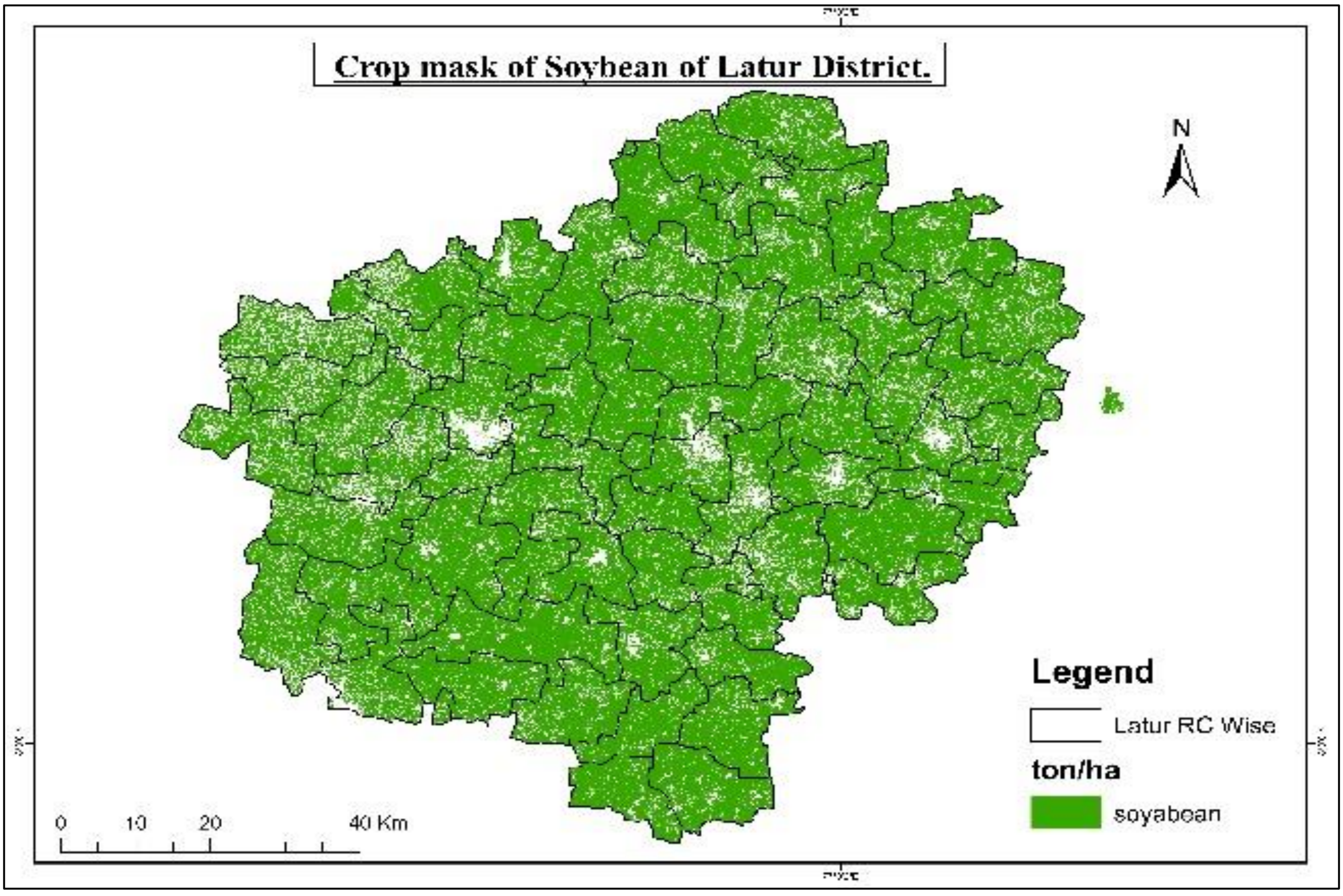

Fig. 7: Soybean Crop Mask of Latur during

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

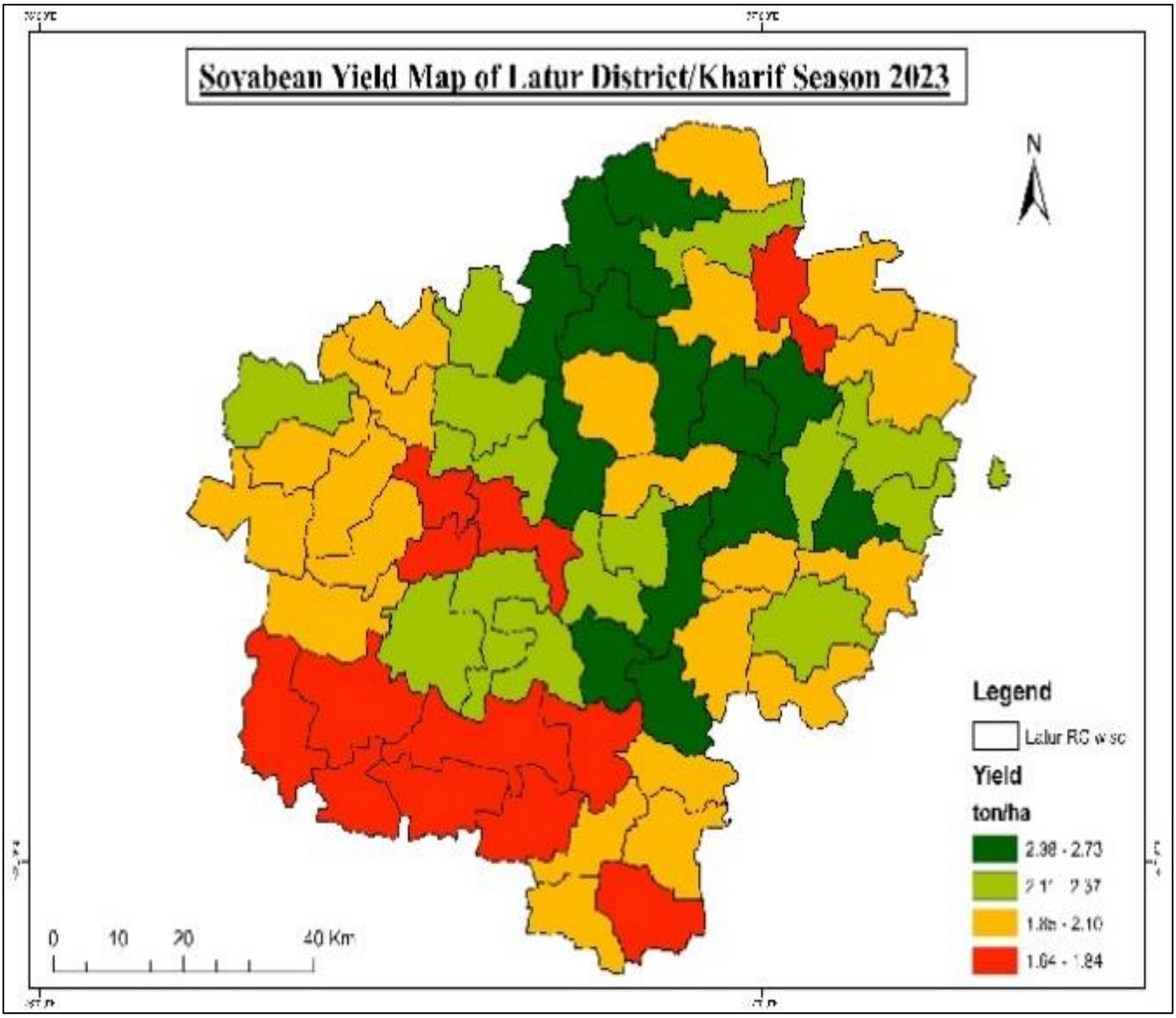

Fig. 8: Soybean Yield of Latur during *Kharif* 2023 *kharif* 2023

- Ahamadpur, Andhori, and Kingaon demonstrated relatively higher actual yields, surpassing 2 tonnes per hectare.
- Hadolati, Belkund, Bhada, Killari, and Shirur Tajband had comparatively lower actual yields, falling below 2 tonnes per hectare.
- Semi-Physical (NPP) Yield also showed variation, with some areas like Ujani, Chakur, and Latur displaying lower performance.
- Notable outliers include Halgara, which had exceptionally high actual yields, and Nalgir, which had a notably high Semi-Physical Yield.
- The dataset suggests a need for closer examination of factors influencing yield discrepancies between regions, such as soil quality and agricultural practices.
- Overall, most regions achieved yields above 2 tonnes per hectare, indicating satisfactory performance for soybean cultivation in 2023.
- Regional variations in climate and agricultural techniques likely influenced the observed differences in soybean yields. Same results were reported by Xiao, X., *et.al* (2006) and Yao, Y., *et.al* (2021)

B. Crop Simulation Model DSSAT-4.8

- The highest yielding RCs are Halgara and Wadhavana (Bk), both with an average yield of 2.89 tonnes per hectare.
- The lowest yielding RCs are Nalgir and Tandulja, with average yields of 0.77 tonnes per hectare and 1.18 tonnes per hectare, respectively.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

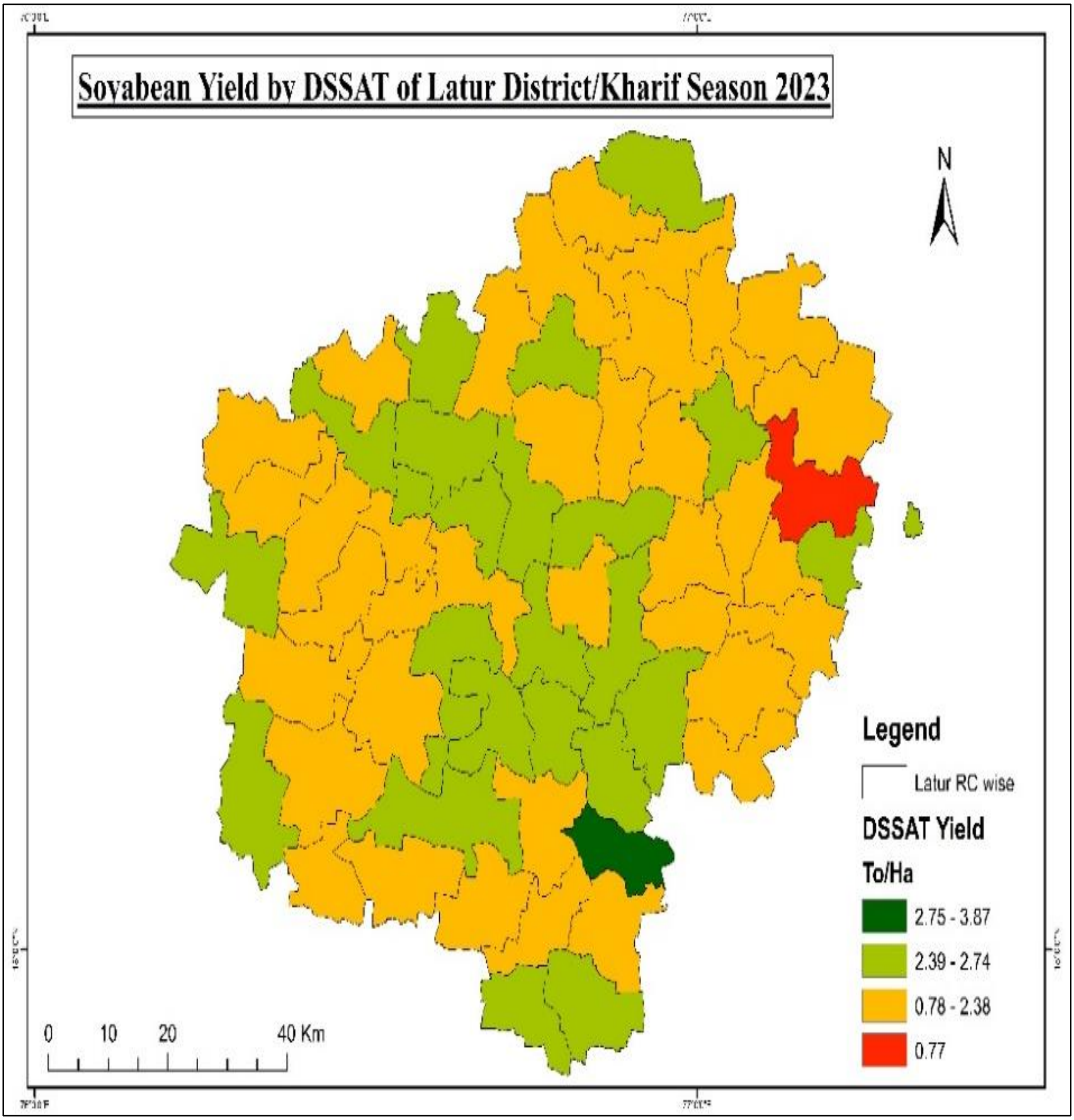

Fig 9: Soybean Yield in T/ha by DSSAT for Latur during *Kharif* 2023

- There is a significant variation in yield between different RCs, with the highest yielding RCs producing more than four times the yield of the lowest yielding RCs.
- There is a significant variation in yield between different RCs, with some RCs having double the yield of others.
- The results suggest the potential benefits of using DSSAT for predicting soybean crop yields, although specific environmental factors and RC conditions may influence the accuracy of the predictions. 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.

C. *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.

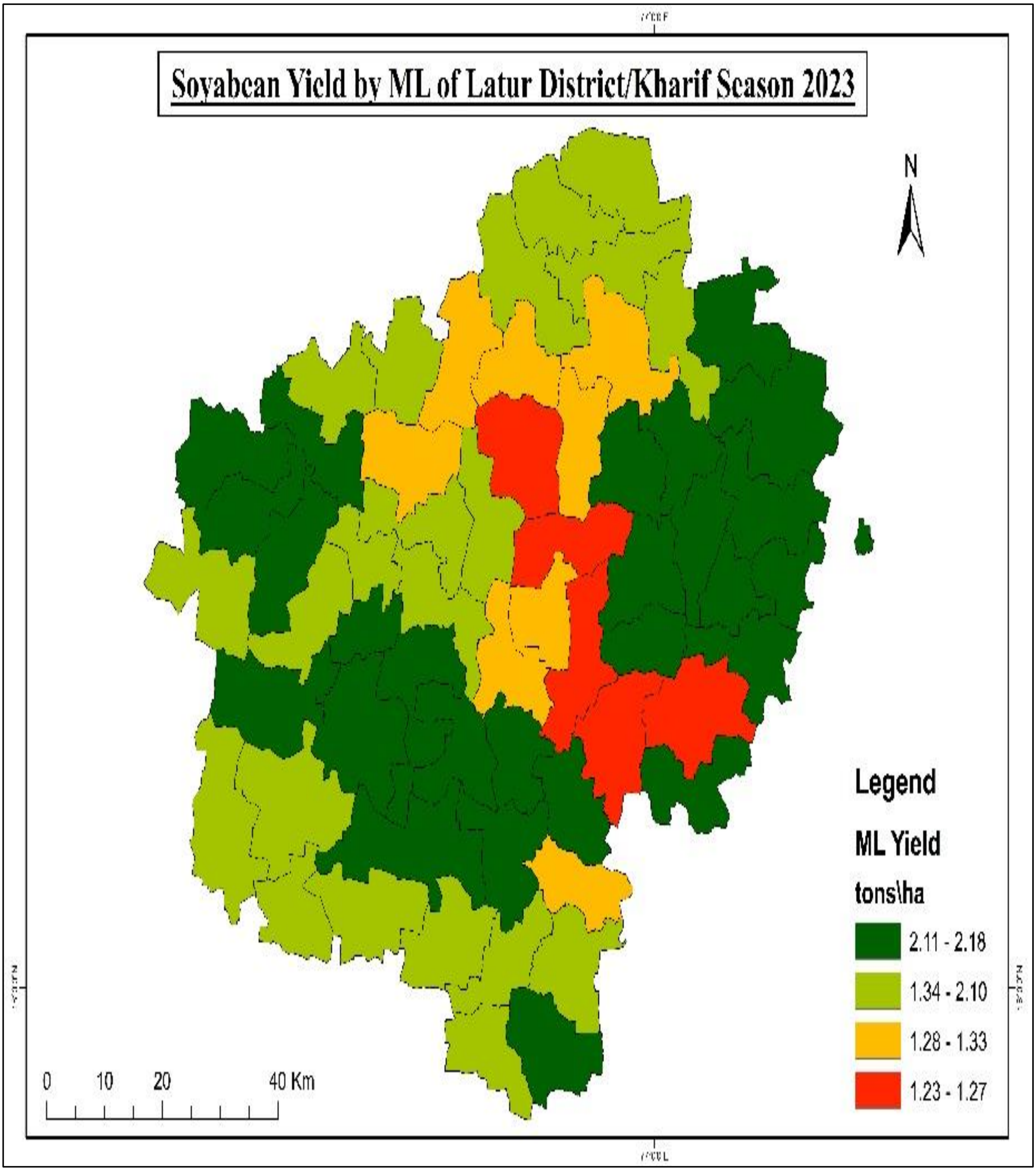

Fig 10: Soybean Yield in T/ha by ML for Latur during *kharif* 2023

- Ahamadpur, Andhori, and Kingaon exhibited relatively stable yields around 2 tonnes per hectare in both actual and ML projections.
- Shirur Tajband, Chakur, and Nalegaon showed lower yields, indicating potential challenges in those regions.
- Halgara stood out with remarkably high actual yield but substantially lower ML yield, suggesting potential discrepancies in the ML model for that area.
- Optimization of ML models can enhance predictive accuracy and contribute to better-informed agricultural practices.
- The results indicate the potential of ML to improve yield predictions and optimize crop management strategies.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

D. Ensemble Model:

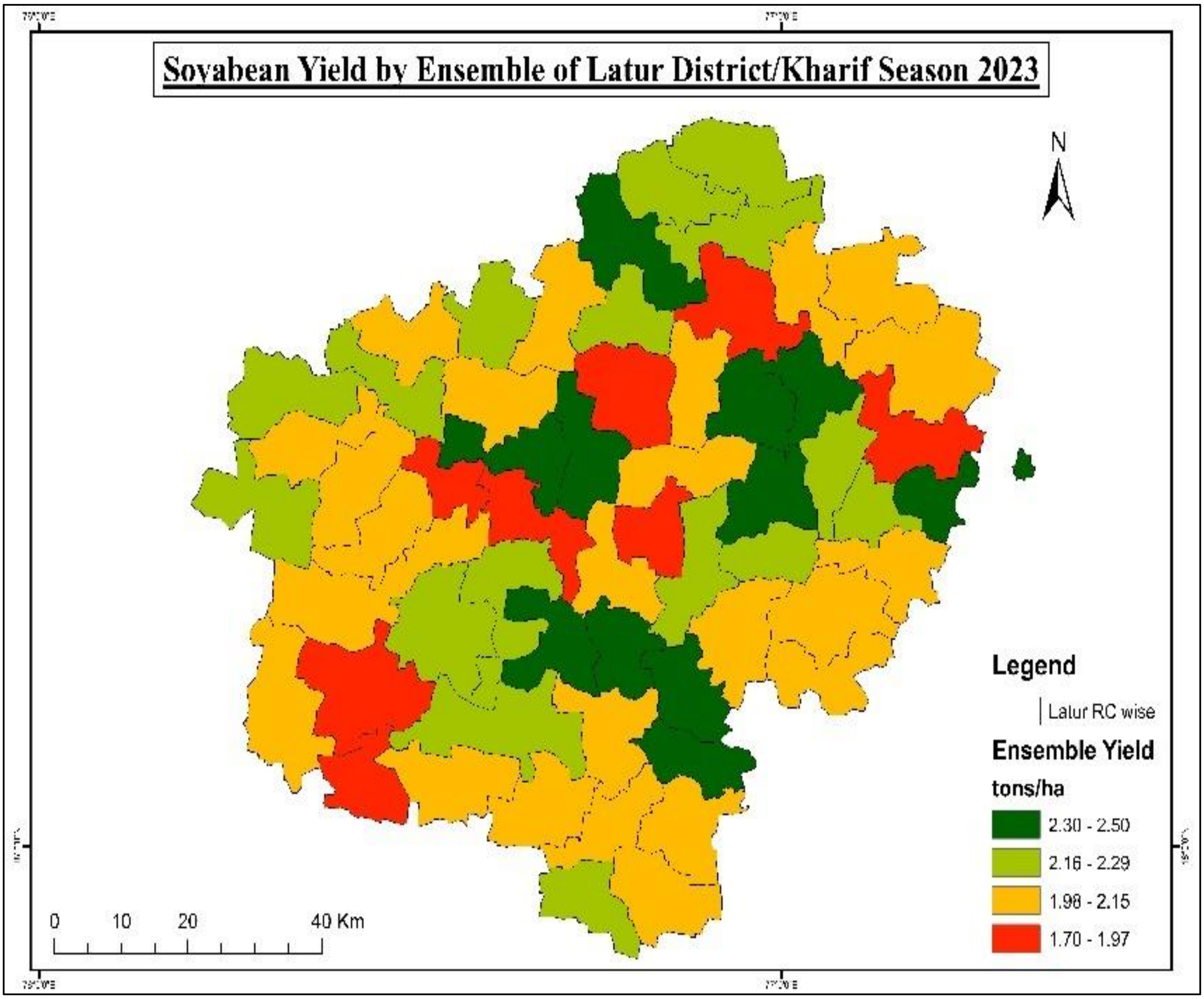

Fig 11: Soybean Yield in T/ha by Ensemble Model for Latur during *Kharif* 2023

Table 4: 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

- The Ensemble Yield represents a combination of all above three predictive models or methods to estimate soybean crop yield.
- Regions like Tandulja, Hisamabad, and Sakol exhibited significant negative percent Error, suggesting considerable discrepancies between actual and predicted yields.
- Conversely, regions like Walandi and Shelgaon showed positive percent Error, indicating slight overestimations in yield predictions.
- Ensemble Yield tended to align closely with Actual yield in some regions, such as Dewrjan and Borol, where the percent Error was minimal or zero.
- The summary reveals the potential of ensemble techniques in predicting soybean yields, though adjustments may be needed to enhance accuracy in regions with high % Error.
- Understanding and minimizing % Error can facilitate better decision-making for farmers and policymakers, optimizing agricultural practices and resource allocation.
- Continuous refinement and validation of predictive models can contribute to more reliable yield forecasts, supporting sustainable soybean production in Latur district. 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.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1981

Table 5: Estimated Yield of Soybean Crop in Tones/Hectors with Different Models and Percent Error with Ensemble Model for Year 2023

District	Tehsil	RC	Field CCE	DSSAT Yield	Semi-Physical Yield	Machine Learning Yield	Ensemble Yield	RMSE % Error
Latur	Ahmadpur	Ahamadpur	1.69	2.31	2.22	2.10	2.22	-31
Latur	Ahmadpur	Andhori	2.14	2.28	2.43	2.06	2.27	-7
Latur	Ahmadpur	Hadolati	1.77	2.30	1.64	2.10	2.01	-14
Latur	Ahmadpur	Khandali	1.58	2.66	2.01	2.08	2.28	-44
Latur	Ahmadpur	Kingaon	2.08	2.33	2.73	2.06	2.40	-15
Latur	Ahmadpur	Shirur tajband	1.35	2.05	2.01	1.33	1.84	-37
Latur	Ausa	Ausa	2.12	2.23	2.29	2.15	2.23	-5
Latur	Ausa	Belkund	1.84	2.18	1.68	2.07	1.97	-7
Latur	Ausa	Bhada	2.19	2.11	1.87	2.15	2.04	7
Latur	Ausa	Killari	2.00	2.27	1.73	2.06	2.02	-1
Latur	Ausa	Kinithot	2.28	2.42	2.14	2.12	2.24	2
Latur	Ausa	Lamjana	2.18	2.53	1.84	2.12	2.18	0
Latur	Ausa	Matola	2.08	2.09	1.72	2.08	1.96	6
Latur	Ausa	Ujani	2.71	2.51	1.70	2.08	2.11	22
Latur	Chakur	Ashta	1.82	2.51	2.42	2.08	2.36	-30
Latur	Chakur	Chakur	1.30	2.21	2.68	1.29	2.13	-63
Latur	Chakur	Nalegaon	1.34	2.60	1.99	1.23	2.02	-51
Latur	Chakur	Shelgaon	2.26	2.32	2.48	2.12	2.32	-3
Latur	Chakur	Wadwal (Na)	1.28	2.26	1.97	1.24	1.89	-47
Latur	Chakur	Zari Bk	1.35	2.74	2.41	1.28	2.24	-66
Latur	Deoni	Borol	2.04	2.06	2.01	2.15	2.07	-2
Latur	Deoni	Deoni	1.41	2.33	2.23	1.25	2.01	-42
Latur	Deoni	Walandi	2.58	2.54	1.98	1.23	1.99	23
Latur	Jalkot	Ghonasi	2.06	2.19	1.94	2.15	2.09	-1
Latur	Jalkot	Jalkot	2.37	2.22	2.05	2.11	2.13	10
Latur	Latur	Babhalgaon	1.57	2.05	1.77	2.06	1.95	-24
Latur	Latur	Chincholi (Bk)	2.26	2.10	2.10	2.14	2.11	7
Latur	Latur	Gategaon	1.93	2.26	2.03	2.13	2.15	-11
Latur	Latur	Harangul (Bk)	1.98	2.10	1.89	2.08	2.02	-2
Latur	Latur	Kanheri	2.01	2.32	1.70	2.11	2.05	-2
Latur	Latur	Kasarkheda	1.97	2.57	2.23	2.07	2.32	-18
Latur	Latur	Latur	2.14	2.14	1.68	2.09	1.97	8
Latur	Latur	Murud (Bk)	2.20	2.53	1.88	2.10	2.19	1
Latur	Latur	Tandulja	1.18	2.33	2.36	2.15	2.29	-94
Latur	Nilanga	Ambulga (Bk)	2.47	2.48	2.47	2.12	2.38	4
Latur	Nilanga	Aurad (Sha)	2.06	2.06	2.05	2.09	2.07	0
Latur	Nilanga	Bhutmugli	1.99	2.11	1.87	2.08	2.02	-1
Latur	Nilanga	Halgara	2.89	3.87	1.91	1.32	2.50	14
Latur	Nilanga	Kasar Balkunda	2.12	2.59	1.65	2.13	2.13	-1
Latur	Nilanga	Kasar Shirashi	2.28	2.65	1.92	2.07	2.24	2
Latur	Nilanga	Madansuri	2.02	2.29	1.75	2.06	2.04	-1
Latur	Nilanga	Nilanga	2.02	2.23	1.82	2.12	2.05	-2
Latur	Nilanga	Nitur	2.53	2.61	2.46	2.14	2.43	4
Latur	Nilanga	Panchincholi	2.39	2.53	2.25	2.14	2.33	3
Latur	Renapur	Karepur	1.23	2.26	2.66	1.30	2.15	-74
Latur	Renapur	Palsi	1.39	2.25	2.00	1.93	2.08	-49
Latur	Renapur	Pangaon	2.18	2.53	2.20	2.09	2.29	-5
Latur	Renapur	Poharegaon	2.30	2.55	2.06	2.18	2.28	1
Latur	Renapur	Renapur	1.39	2.65	2.11	1.31	2.11	-52
Latur	Shirur-Anantpal	Hisamabad	1.09	2.45	2.15	1.29	2.04	-86
Latur	Shirur-Anantpal	Sakol	1.32	2.66	2.52	1.27	2.24	-70
Latur	Shirur-Anantpal	Shirur Anantpal	1.53	2.06	2.18	1.29	1.90	-24
Latur	Udgir	Dewrjan	2.16	2.38	1.93	2.15	2.16	0
Latur	Udgir	Her	2.39	2.30	2.48	2.15	2.32	3
Latur	Udgir	Mogha	2.12	2.22	2.03	2.17	2.14	-1
Latur	Udgir	Nagalgaon	2.40	2.59	2.20	2.15	2.34	2
Latur	Udgir	Nalgir	1.57	0.77	2.37	2.15	1.70	-8
Latur	Udgir	Tondar	2.21	2.08	2.33	2.18	2.20	0
Latur	Udgir	Udgir	2.36	2.19	2.53	2.12	2.29	3

Latur	Udgir	Wadhavana(Bk)	2.62	2.56	2.68	2.12	2.49	5
Latur district Average values =			**1.968**	**2.342**	**2.107**	**1.929**	**2.149**	**-14.41**

Table 6: Average Percent Error of all Approaches Estimated Yield of Soybean Crop for Year 2023.

Methods	Field CCE	DSSAT Yield	Semi-Physical Yield	Machine Learning Yield	Ensemble Yield
Yield (T/h)	1.97	2.34	2.11	1.93	2.15
RMSE % Error		-19	-7	2	-9

In Table 4, the yield estimated by various methods is presented, including the percentage error of yield by the ensemble model with field CCE, which is provided in the last column. Out of 62 points only 14 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. While, in table no. 5 average yields and % errors for whole Latur district are mentioned. All values of % error presenting error was less than ±30%. Which indicates that the process adopted for RC wise Soybean yield estimation is acceptable for Latur district for all the models.

IV. CONCLUSION

"The comparative evaluation of NPP, DSSAT, and Machine Learning models for predicting soybean crop yields in Latur, 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 Latur District of Maharashtra State."

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.