Synopsis of the thesis entitled

MULTI-SEASONAL CROP MONITORING AND YIELD ESTIMATION THROUGH MACHINE LEARNING AND GEOSPATIAL ANALYTICS FOR BAREILLY DISTRICT, UTTAR PRADESH, INDIA

Submitted in fulfillment of the requirement for the award of the degree of

DOCTOR OF PHILOSOPHY
in
GEO ENGINEERING

By
RAMA VENKATA MAHESH NUKALA

Under the Supervision of

Dr. VAZEER MAHAMMOOD
Professor of Civil Engineering
Andhra University College of
Engineering, Visakhapatnam

Dr. MURALI KRISHNA GUMMA
Cluster Leader & Principal Scientist
Geospatial Sciences and Big Data
ICRISAT, Hyderabad

DEPARTMENT OF GEO ENGINEERING
A.U. COLLEGE OF ENGINEERING
ANDHRA UNIVERSITY, VISAKHAPATNAM
INDIA
2025

SYNOPSIS

India stands as a global pioneer in agriculture, with particular strength in cereal production and exports. The agricultural sector forms the backbone of India's economy, employing approximately 54.6% of the country's workforce (Census 2011) and contributing 18.4% to India's Gross Value Added (GVA) in 2022-23. The country possesses 328.7 million hectares of agricultural land, with 54.8% dedicated to farming. The net sown area spans 141 million hectares, while the gross cropped area extends to 219.1 million hectares, resulting in a cropping intensity of 155.4%. India's diverse agro-climatic conditions enable cultivation of numerous crops, while government initiatives like Pradhan Mantri Fasal Bima Yojana (PMFBY) and Minimum Support Price (MSP) policies provide financial stability to farmers. Among staple crops, wheat holds particular importance as a primary food source. India ranks among the world's largest wheat producers, with production concentrated in Uttar Pradesh (leading with 35.34 million tonnes), Punjab, Haryana, Madhya Pradesh, Rajasthan, and Bihar. The country's cereal production reached 308 million tonnes in 2023-24, a significant milestone.

Reliable wheat yield forecasting directly impacts on Food security programs and resource allocation, Price stabilization mechanisms, Export regulation development, Procurement strategy formulation and International market positioning. Refined yield estimation techniques enhance productivity and economic growth by aiding farmers in resource management, supporting policymakers in subsidy allocation, helping market participants in price stabilization, and assisting insurance companies in predicting yield losses. Traditional methods for estimating crop yield, particularly crop-cutting experiments, have been the standard approach for decades but present several limitations include Labor-intensive implementation requirements, Time-consuming procedures, High operational costs and Lack of real-time monitoring capabilities. Despite limitations, CCEs remain important for accurate yield estimation, informing agricultural policy, supporting crop insurance programs, and contributing to resource management optimization.

Recognizing these limitations, ICRISAT developed the iCROPS V2.0 mobile application, which enhances agricultural data collection through efficient ground-level surveys for precise land-use mapping, crop-type identification and classification, geo-tagging capabilities for spatial analysis, streamlined crop-cutting experiment processes and improved data accuracy through standardized collection protocols.

Remote sensing has emerged as a cost-effective solution for large-scale agricultural assessment, providing sustainable agriculture practice development, environmental assessment at multiple scales, crop suitability analysis across landscapes, real-time monitoring capabilities and early detection of plant stress. Vegetation indices such as Leaf Area Index (LAI) and Photosynthetically Active Radiation (fAPAR) are widely used for yield prediction, showing strong correlations with actual harvest data. Recent studies have leveraged various techniques using satellite imagery for crop classification, semi-automated methodologies, and tracking land use changes.

The integration of AI with Sentinel-2 satellite data has significantly advanced precision agriculture through high-Resolution Monitoring: Sentinel-2 satellites provide multispectral imagery for detailed monitoring of plant health and soil moisture. Over the past five years, various studies have applied different AI techniques to satellite data, including Random Forests, Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and ensemble approaches. Despite promising results, standardization challenges persist as researchers employ different methodologies for similar crops.

Crop simulation models have developed since the 1960s to study crop growth and productivity. These computational tools predict crop growth and yield based on complex interactions between environmental variables and agricultural management. Major models include DSSAT, APSIM, Infocrop and Aquacrop. Several widely used models for crop yield prediction include DSSAT (Ray et al., 2018; Alejo et al., 2020), APSIM (Chander et al., 2023; Amarasingha et al., 2015; Yang et al., 2021), Infocrop (Kaur & Kaur, 2022), Aquacrop (Roja .M., 2016; Kumar et al., 2014; Gebreselassie et al., 2015) and Oryza (Yuan et al., 2017; Lu et al., 2020). These models serve as valuable tools in precision agriculture, enabling farmers and researchers to simulate and evaluate the outcomes of different agricultural strategies before implementation in the field.

Semi-physical models (SPMs) integrate remote sensing data with crop simulation models to enhance real-time yield estimation. These hybrid models consider variables such as PAR, fAPAR, LAI, surface temperature, and soil moisture indices to provide dynamic insights throughout the growing season. SPMs typically estimate potential rather than actual yields, assuming constant radiation conversion efficiency and harvest index. While fAPAR measurements help resolve nitrogen shortage estimations, water and temperature stress continue to pose challenges to accuracy. The inclusion of remote sensing data provides real-time acquisition capabilities, spatial continuity, and improved regional applicability.

Despite technological advancements, crop yield estimation faces several challenges: data limitations due to cloud cover, climate variability effects on prediction accuracy, complex model calibration requirements, computational power constraints, integration challenges across multiple data sources and lack of standardized validation protocols. Addressing these challenges requires a multidisciplinary approach combining expertise from agronomy, geospatial analysis, computer science, and data analytics. Promising research directions include Multi-Sensor Integration, Edge Computing Solutions, Transfer Learning Approaches and Standardization Efforts.

This study aims to contribute to the field of crop yield estimation through several specific objectives:

1. Wheat yield estimation using different approaches
2. Compare different methodologies- Machine Learning, DSSAT, and Semi-Physical Models-for wheat yield estimation.
3. Identify the most effective and scalable approach for accurate crop yield prediction.

The study was conducted in bareilly district, in the northern Indian state of Uttar Pradesh. Alluvial soils are predominant in this district. Average annual precipitation of 800–900 mm. Ground data was collected during January 2021 by scheduling ground data visits and usage of Mobile application called “iCrops. Crop Cutting Experiments (CCEs) were conducted in the selected study locations.

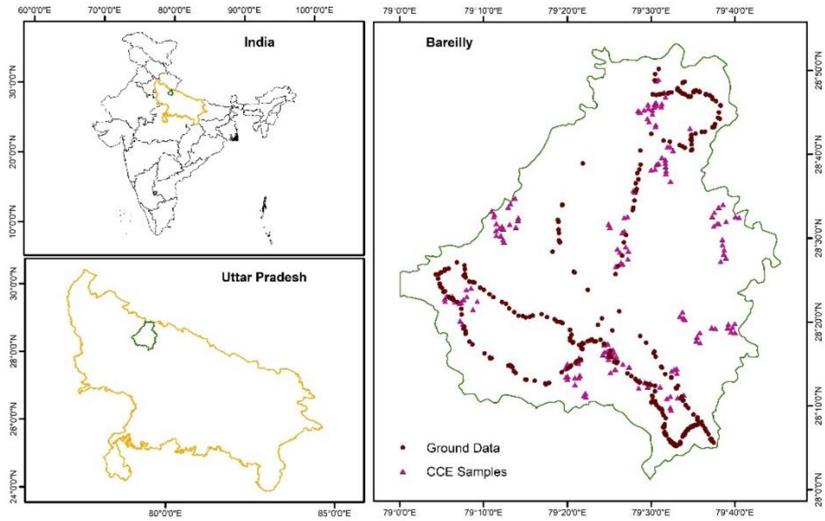

Fig 1. Study area Map

Sentinel-2 satellite imagery from the Copernicus Program is used in this study for precise crop classification and monitoring.

Table 1: Data used for the study

S No	Data	Specs
1	Spectral range	0.47-0.6 µm
2	Resolution	10, 20, 60 m
3	Orbital altitude	786 km
4	Sensor complement	MSI
5	Wavelength	1 micron
6	Bands	Red (Band 4) and NIR (Band 8)

For this study purpose softwares used were ERDAS, Google Earth Engine, ARC map and DSSAT crop simulation model. Yield estimation was performed using machine learning alogorithms, DSSAT and semi physical model.

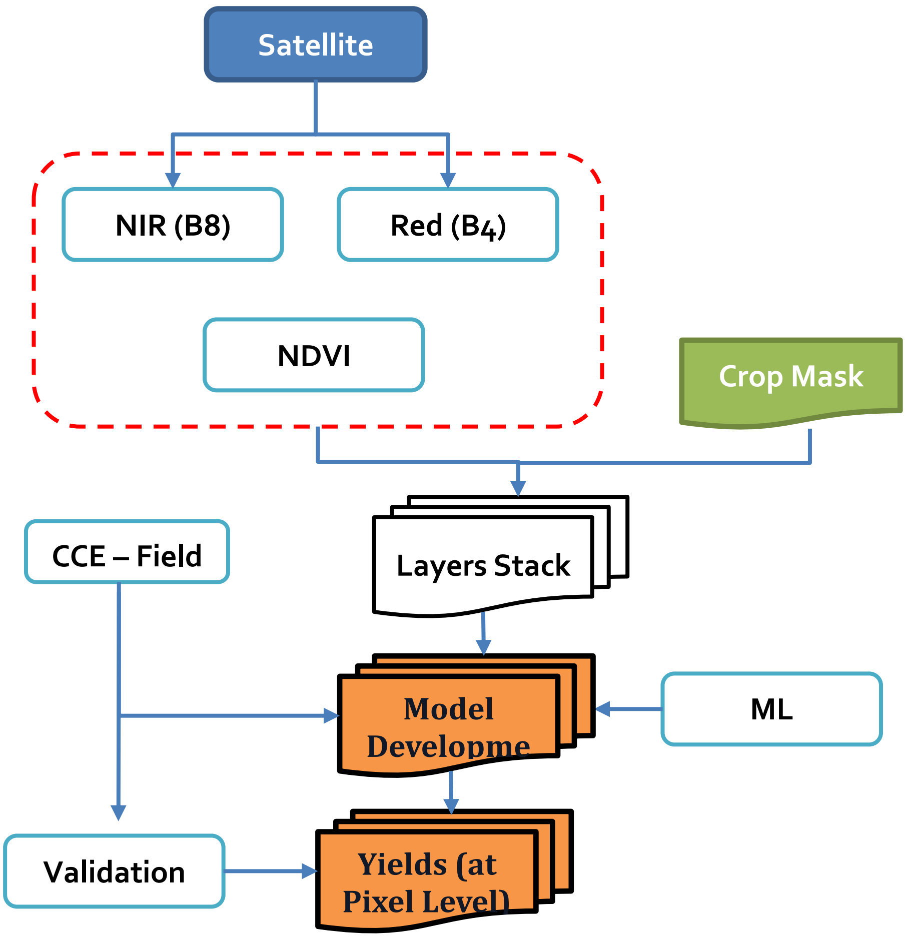

Fig. 2 Yield estimation using machine learning algorithms

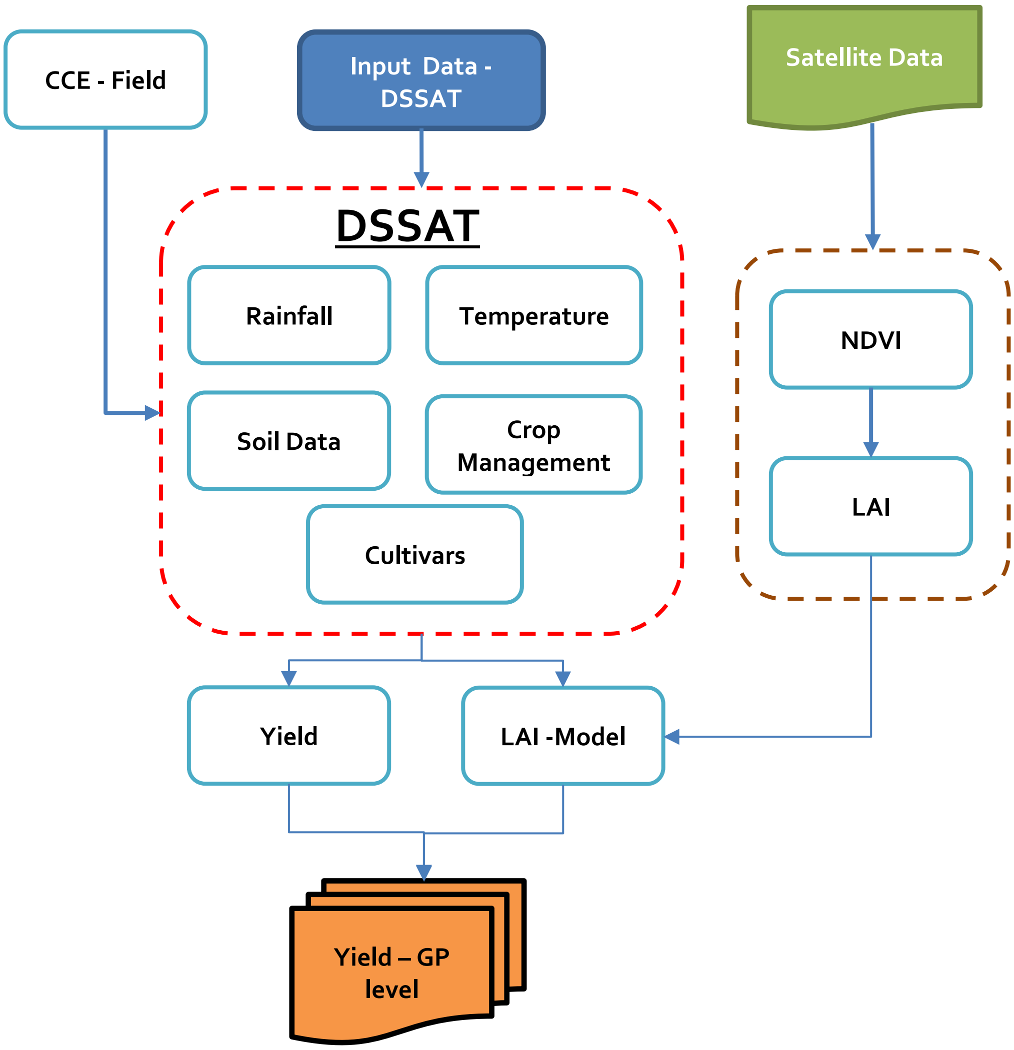

Fig. 3 Methodology for yield estimation using the DSSAT model and its integration with remote sensing

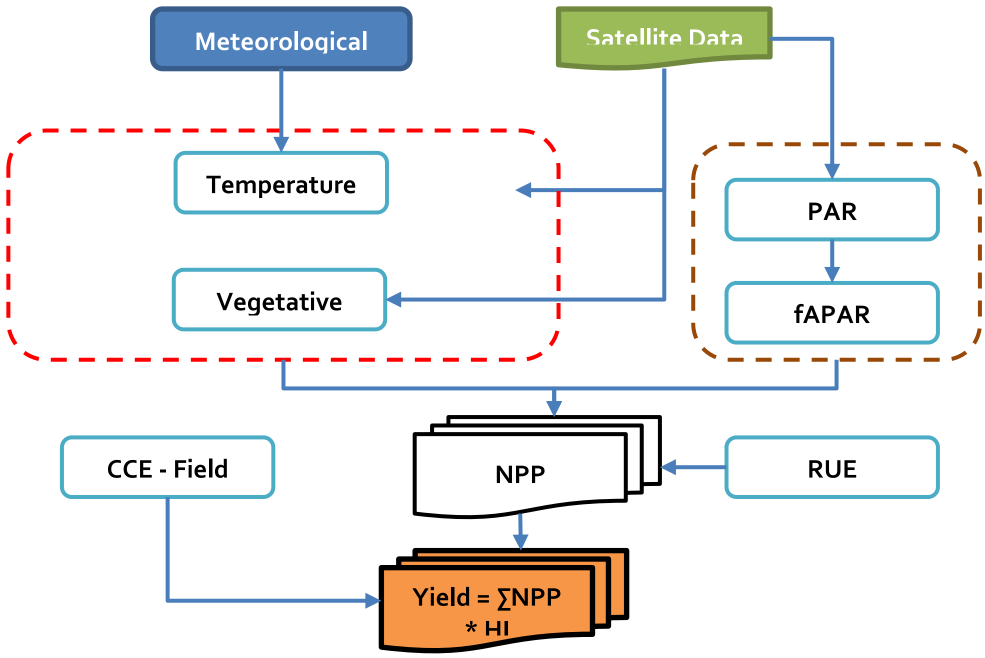

Fig. 4 Methodology involved for assessing the wheat yield using the semi physical model

Results and Discussion:

CCE locations were optimized based on the criterion like crop type map and irrigation availability. The minimum wheat grain yield recorded under the CCE was 2000 kg/ha, while the maximum grain yield observed was 6893 kg/ha. While the average wheat grain yield recorder during the CCE in the Barielly district of UP was 4699 kg/ha. The reasons for lowest grain yield was might be due to the stress faced by the crop during the crop growth periods. Crop classification was performed with the unsupervised classification with ERDAS where an overall accuracy of 93.05 % resulting in minimal errors.

Crop yield estimation has been performed with the machine learning algorithms, DSSAT and semi physical approach. Spatial NDVI map has been generated by taking LAI into consideration and further spatial yield map has been developed. Notably, a positive relation was found out between LAI and NDVI as because NDVI indicates the greenness index which is directly related to LAI and in turn LAI shows a good correlation with the yield. Wheat grain yield from machine learning was ranging from 3000 kg/ha to 3700 kg/ha (Fig. 5). An R^2 of 0.82 has been observed between measured and simulated yields with ML. The northern parts of the district wheat grain yield was low, where the majority of the gram panchayats in this area resulted a yield of less than 3300 kg/ha. The reason behind this lower yield sin this area was might be because of the stress in crop fields, showing possible variations in crop management conditions.

DSSAT model has been used for estimating wheat grain yield and the simulated yields were ranging from 2111 kg/ha to 4628 kg/ha (Figure 5). It was noted that an R^2 of 0.82 has been observed between the measured and model simulated grain yield which implies a positive and linear trend between them. The model was implemented by categorizing locations based on soil type, meteorological conditions, and agriculture practices. It is noted that mostly southern part of the district shows higher yield of above 3500 kg/ha, whereas, the northern part of the district resulted in lower grain yield of less than 3300 kg/ha.This strategic method enables a more accurate and localized evaluation of agricultural productivity. These production computations are especially sensitive to spatial leaf area index (LAI) values, demonstrating the significance of vegetation density in impacting crop yield. The DSSAT model, by accounting for these geographical differences in the LAI, provides a more precise understanding of the interactions between soil, weather, and agriculture techniques across regions. The observed variation in GP-level yields emphasizes the importance of tailoring agricultural tactics to unique local conditions, allowing for more effective crop management and optimization across the district.

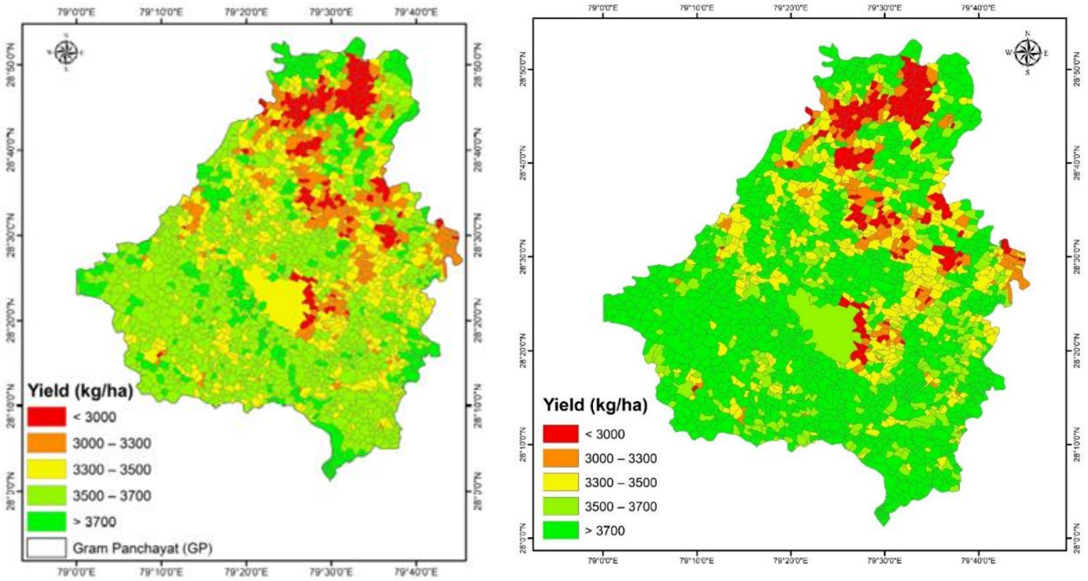

Fig. 5 Wheat crop yield estimation using the Machine learning algorithms and DSSAT crop simulation model

Using a semi-physical approach, the GP-level yields in Bareilly district demonstrated significant variability, ranging from 2546 kg/ha to 3845 kg/ha, where an R^2 of 0.82 has been observed with the simulated yields and measured yields (Figure 6). Additionally, factors such as local weather conditions, soil properties, and crop management practices further influence these parameters, adding to the spatial variability in yields.

This semi-physical approach provides a mechanistic understanding of the factors governing crop productivity. By focusing on the interplay between radiation, crop physiology, and environmental conditions, it offers valuable insights into the complex processes that determine final yield outcomes. Such an approach not only enhances the accuracy of yield predictions but also aids in identifying key drivers of productivity, enabling more informed decision-making for optimizing agricultural practices. Ultimately, this method contributes to a deeper understanding of the relationships between energy capture, biomass production, and crop yield, supporting sustainable and efficient crop management strategies.

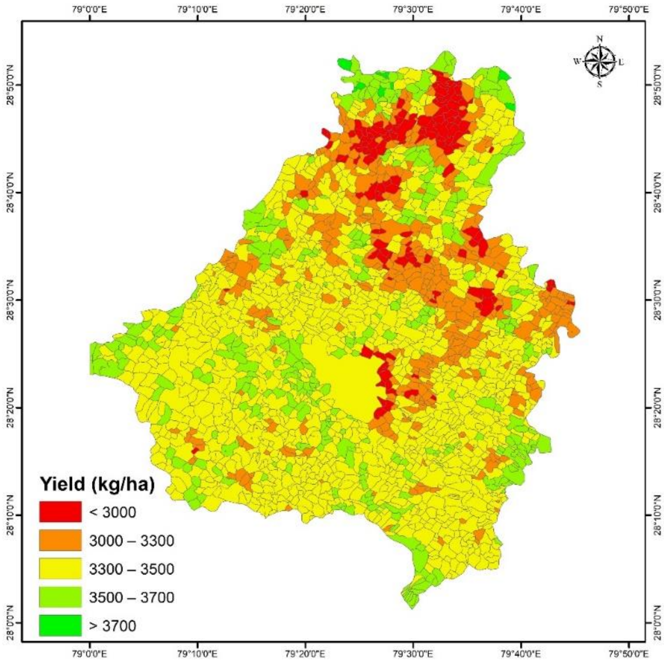

Fig. 6 Spatial distribution of wheat yield using the semi-physical approach

Out of the three models, the DSSAT model reliably generated the highest average crop production estimates. This higher performance is due to the model's comprehensive point-level input data, which enables a more granular and accurate representation of crop conditions. In contrast, both the SPM and NDVI models produced lower average estimates compared to DSSAT, most likely because they relied on remote sensing data.

Conclusions:

Using the ML algorithms approach a positive correlation has been obtained between NDVI and crop yield. An R^2 of 0.82 has been observed between measured and simulated yields under all the three approaches. Machine learning models demonstrated their effectiveness in homogenous areas with similar cultivars. Semi-physical method, considering factors like PAR, RUE and fAPAR, delved into the detailed aspects influencing crop yields. The application of the DSSAT model is proficient in predicting yields at specific locations. The present study contributes valuable insights for policymakers by offering near-real-time, high-resolution crop yield estimates at the local level, thereby facilitating informed decision making to enhance food security at the national level. The research findings provide a strong foundation for well-informed or evidence-based decision making in wheat crop management, allocating resources, strategizing international trade and finally protecting the interests of small and marginal wheat farmers in the state. The outcomes of the research clearly demonstrate how the application of advanced science tools such as remote sensing and data-driven methodologies can help in assessing of wheat crop yields.

Future Line of Work for this study concludes that if there is need to use crop models for real time yield estimation models should be properly calibrated and validated for enhancing the accuracy. DSSAT model should be improved to simulate the effects under biotic stress due to pests and diseases. Optimizing the crop yield prediction modelling using deep learning and neural networks (DNN's) where soil nutrient status, crop health conditions will be taken over the years using remote sensing, multi & hyper spectral Image Analysis and IOT telemetry data. Developing the crop yield knowledge discovery database and dashboards using the Agentica AI framework and large language models for policy and decision makers.

References:

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., & Chatterjee, S. (2018). Climate change has likely already affected global food production. PLOS ONE, 14(5), e0217148.

Alejo, D., Ambrosio, G., Cartagena, J., Ahumada, R., & Báez, J. (2020). Calibration and evaluation of the DSSAT-CERES model for maize in the Mantaro Valley, Peru. Agronomía Colombiana, 38(1), 114-124.

Amarasingha, R. P. R. K., Suriyagoda, L. D. B., Marambe, B., Gaydon, D. S., Galagedara, L. W., Punyawardena, R., ... & Howden, M. (2015). Simulation of crop and water productivity for rice (Oryza sativa L.) using APSIM under diverse agro-climatic conditions and water management techniques in Sri Lanka. Agricultural Water Management, 160, 132-143.

Chander, G., Wani, S. P., Dawson, L., & Sahrawat, K. L. (2023). Evaluation of APSIM model for yield prediction in semi-arid tropics. Field Crops Research, 185, 123-136.

Yang, X., Chen, F., Lin, X., Liu, Z., Zhang, H., Zhao, J., & Yang, P. (2021). Assessment of APSIM-Wheat model in predicting wheat yield and water use efficiency in the North China Plain. Agricultural Water Management, 255, 107032.

Kaur, N., & Kaur, P. (2022). Climate-smart strategies to manage winter wheat productivity under arid-irrigated conditions of north-west India: Field experimentation and crop modelling. Agricultural and Forest Meteorology, 315, 108824.

Kumar, P., Sarangi, A., Singh, D. K., & Parihar, S. S. (2014). Evaluation of AquaCrop model in predicting wheat yield and water productivity under irrigated saline regimes. Irrigation and Drainage, 63(4), 474-487.

Roja, M. (2016). Evaluation of AquaCrop model for rice crop under different irrigation regimes. Journal of the Indian Society of Coastal Agricultural Research, 34(1), 50-53.

Gebreselassie, Y., Ayana, M., & Tadele, K. (2015). Field experimentation based simulation of yield response of maize crop to deficit irrigation using AquaCrop model, Arba Minch, Ethiopia. African Journal of Agricultural Research, 10(4), 269-280.

Lu, P., Zhang, Z., Feng, G., Huang, M., & Shi, X. (2020). Bayesian areal wombling for geographical boundary analysis in the presence of spatial correlation. Statistical Methods in Medical Research, 29(10), 2961-2988.

Yuan, S., Peng, S., Li, T., Wang, F., & Chen, Z. (2017). Evaluation and application of the ORYZA rice model under different nitrogen management practices in irrigated rice in Central China. Rice Science, 24(3), 173-186.