Journal of Information Systems Engineering and Management
2025, 10(13s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Data-Driven Prediction of Nitrogen Stress and Crop Yield in a Maize -Wheat CroppingSystem

Halley Okasa 1b, Shreya 1c, Shilpi Verma 2d, Greeshma Arya 1a,AnchalDass 2a Ashish Bagwari 3e, Ciro Rodriguez 4f,Carlos Navarro 5g

1IndiraGandhiDelhiTechnicalUniversityforWomen,KashmereGate,NewDelhi-110006,India

2ICAR-IndianAgriculturalResearchInstitute,PusaCampus,NewDelhi-110012, India

3Women Institute of Technology, VMSBUTU, Dehradun-248007, India

4,5Universidad Nacional Mayor de San Marcos (UNMSM), Lima15081,Peru

Correspondence author:

a)greeshmaarya@igdtuw.ac.in; anchal_d@rediffmail.com; b)halley087btece20@igdtuw.ac.in, c)shreya169btece20@igdtuw.ac.in

d)shilpi.iari@gmail.com, e)ashishbagwari@wit.ac.in; f) crodriguezro@unmsm.edu.pe; g) cnavarrod@unmsm.edu.pe

ARTICLE INFO

Received: 30 Nov 2024

Revised: 12 Jan 2025

Accepted: 30 Jan 2025

ABSTRACT

This study uses predictive analytics to deal with nitrogen stress and get the best crop returns in a system that grows both wheat and corn. We measured the amount of nitrogen (N) in the leaves along with a number of important farming factors using modern plant monitors like the GreenSeeker and SPAD meter. Different field studies gave us a range of nitrogen treatment rates for wheat (0 to 240 kg N/ha) and maize (40 to 300 kg N/ha). Machine learning methods, like Random Forest and Support Vector Machines, were used to create models that can accurately predict crop yields based on nitrogen doses, Normalised Difference Vegetation Index (NDVI) values, SPAD values, and direct measurements of leaf-N content. The results from pair plot studies showed strong positive links between leaf-N content, NDVI, SPAD values, and the rates of N application. These connections support the accuracy of the prediction models. Notably, the model that was tuned for wheat showed that it could be used with other crops by correctly predicting corn yields. It was amazing how well the prediction models worked. Random Forest and Support Vector Machines were able to get 85.71% and 100% accuracy on the original dataset, respectively.

Keywords: Precision Agriculture, Nitrogen Management, Machine Learning, Crop Yield Prediction, NDVI, SPAD Meter.

1INTRODUCTION

Growing worldwide demand for agricultural output among limited land and water resources calls for the use of sophisticated technology to improve crop management techniques. Crucially important for crop production, nitrogen (N) is responsible for plant development and output. But since nitrogen is dynamic in the ecosystem and has a significant effect on environmental sustainability, good management of it is difficult. By means of focused nutrient management, precision agriculture has become a transforming tool allowing farmers to maximise yields while reducing environmental effects. In this regard, especially in common cropping systems like maize and wheat, the use of predictive analytics in agriculture presents a hopeful route to maximise nitrogen use and raise crop yields. Two basic crops that greatly contribute to the world's food supply are maize and wheat, hence their farming is rather vital for food security. Achieving ideal nitrogen management in these crops is difficult, however, impacted by soil type, climate, and crop phenological phases. Conventional methods of nitrogen delivery are often broad and may not meet the particular requirements of the crops at various development phases, therefore causing inefficiency and environmental damage via leaching and emissions. Recent developments in sensor technology and machine learning have opened the path for more exact and data-driven methods of crop management, therefore addressing these difficulties. This work integrates many important characteristics like the Normalised Difference Vegetation Index (NDVI) and SPAD values to directly evaluate the leaf-N content in maize and wheat crops using GreenSeeker and SPAD meter sensors. These sensors enable real-time nitrogen application decisions by offering fast, non-destructive, precise readings of plant health and nitrogen status [1]. This work intends to build strong prediction models that can estimate crop yield and nitrogen needs with great accuracy by linking these sensor values with nitrogen application rates (range from 0 to 240 kg N/ha in wheat and 40 to 300 kg N/ha in maize). Leveraging machine learning models—more especially, Random Forest and Support Vector Machines (SVM)—improves the forecast accuracy of nitrogen influence on agricultural yields. These models include variables like nitrogen dosages, NDVI, SPAD values, and leaf-N content and are trained on large sets gathered from field studies. Deeper understanding of the interplay between nitrogen treatment and crop performance results from

these models' capacity to learn complicated linkages between many variables, hence surpassing conventional statistical approaches [2].
By lowering the danger of nitrogen misuse and related environmental effects, the combined method of combining sensor technologies with machine learning not only improves nitrogen use efficiency but also promotes sustainable farming practices [3]. Moreover, the cross-valuation of the wheat model on maize datasets emphasises the adaptability and flexibility of the built predictive models, implying that same strategies might be effectively implemented to different crops and surroundings. By proving how predictive analytics may be properly used to control nitrogen stress and improve crop yields in maize-wheat cropping systems, this study generally advances the discipline of precision agriculture. Farmers may improve crop production, maximise input use, and lower environmental consequences by switching to more data-driven, exact agricultural methods, therefore aiding worldwide attempts towards sustainable agriculture. This work not only deepens our knowledge of crop-nitrogen dynamics but also offers useful information that can be used in actual agricultural environments to raise the sustainability and efficiency of crop output.

2. RELATED WORK

Recent research has examined closely how predictive analytics may be used into agricultural operations, especially to control nitrogen stress and improve crop yields. Previous studies have shown the great influence of exact nitrogen management on the production of important crops such maize and wheat, so stressing the need of advanced tools and techniques to maximise nitrogen use efficiency and lower environmental risks related with too strong nitrogen application [4][5]. Development and implementation of sensor technologies constitute a fundamental feature of this study field. Measurement of crop vigour and nitrogen content straight from the field has made growing use of instruments as the GreenSeeker and SPAD meters. Making wise nitrogen management choices depends on reliable and real-time insights into plant health, which studies by researchers [6][7] have shown these sensors may provide. These results complement our method by stressing the pragmatic relevance of such sensors in nitrogen management forecasting models.

Recent studies aiming at improving agricultural decision-making have also concentrated mostly on machine learning models. Random Forest and Support Vector Machines, for example, have been extensively embraced because of their capacity to manage vast datasets with various input factors and their resilience in catching complicated nonlinear correlations between these variables [8][9]. From controlling irrigation schedules to predicting insect outbreaks, these models have shown success in many predictive activities in agriculture. Reflecting a growing trend in the literature, our work expands this line of research by using these models to forecast crop yields and nitrogen demands in maize-wheat cropping systems [10][11]. Furthermore, the research shows a good association between NDVI values, SPAD measures, and nitrogen levels, therefore verifying the accuracy of these indices as markers of crop condition [12]. This complements our approach, in which our machine learning models depend critically on NDVI and SPAD measurements. Our pair plot studies show that the favourable correlation between these indices and nitrogen application rates offers further evidence for using these measures in predictive analytics systems for crop management.
Research on the scalability and transferability of machine learning models throughout many crops and habitats has also been conducted [13]. This feature is especially pertinent to our research as a model developed on wheat data is also evaluated on maize, proving its general relevance throughout crop kinds. This cross-crop validation indicates the possible general relevance of our prediction models and implies that comparable approaches may be successful in other agricultural environments. The corpus of related work supports the method used in our research and emphasises the need of incorporating sensor data and machine learning methods into contemporary agricultural operations. Particularly in relation to nitrogen stress management and crop yield optimisation in maize-wheat cropping systems, our study helps to advance the discipline of precision agriculture by expanding on the current knowledge and approaches recorded in the literature. The consistent results of many research highlight the possibility of these technologies to change agricultural operations, therefore producing more environmentally friendly and profitable farming solutions.

Table 1: Related Work Summary

Technology Used	Key Parameter	Finding	Crop Focused
GreenSeeker	NDVI	Accurate real-time plant health monitoring enhances nitrogen management decisions.	Wheat, Maize
SPAD Meter	Chlorophyll content	Provides reliable measurements of leaf nitrogen content, influencing fertilization strategies.	Wheat, Maize
Random Forest	Multiple variables	High predictive accuracy in yield forecasting based on comprehensive input data.	Wheat
Support Vector Machines	Multiple variables	Demonstrated 100% accuracy on the dataset, showcasing robust performance in yield prediction.	Maize

NDVI Sensors	NDVI	Positive correlation with nitrogen levels, enabling better nutrient management.	General Crops
Machine Learning Models	Leaf-N content	Effective in predicting nitrogen needs and optimizing yield outcomes.	Wheat, Maize
SPAD Meters	SPAD values	Correlated with nitrogen application rates, useful for adjusting nitrogen inputs.	Wheat, Maize
Predictive Models	Crop health metrics	Validated across different crops, showing adaptability of models for varied agricultural uses.	Wheat, Maize
Sensor Technology	Plant health indices	Enhances precision agriculture by providing data for real-time decision-making.	General Crops
Cross-Crop Validation	Model accuracy	Models trained on one crop type successfully predict outcomes for another, increasing utility.	Wheat, Maize
Real-Time Monitoring	Crop vigor	Supports sustainable agriculture practices by optimizing input use and reducing waste.	General Crops
Environmental Impact Studies	Nitrogen usage	Reduces risks associated with nitrogen overuse through targeted management practices.	General Crops

3. MATERIALS AND METHODS

The approaches of the research are very crucial for understanding how machine learning might be used with agricultural techniques to forecast food output when nitrogen levels vary. The ICAR-IARI facility in New Delhi, with a warm and semi-arid environment, hosted the experiments. This made it ideal for researching the wheat and maize products throughout the kharif and rabi seasons. Careful planning went into the field testing to see how nitrogen fertilisation influenced crop development. For both crops, they used a fixed plot design with similar field patterns. Each crop experiment had 21 plots, and seven different nitrogen treatment rates were used in each plot three times. This organised method made it possible to compare things with different amounts of nitrogen in a safe way. Standard farming methods were used on all plots, such as applying important nutrients like phosphorus (P), potassium (K), and zinc (Zn) at the same time every time and watering at the right times. Herbicides that were recommended and hand pulling of weeds were used to get rid of them.

Nitrogen was added to wheat at three important growth stages: planting, tillering, and booting. Applying it to corn at the planting, knee-high, and pre-tasseling stages was the right thing to do. This split spraying method is meant to help plants absorb nitrogen better while reducing losses through washing or volatilisation. Important agricultural data were gathered, such as the amount of nitrogen in the leaves, the NDVI and SPAD values, the dry weight of the plants, the grain yield, and the straw output. The alkaline permanganate method, a solid precise chemistry method, was used to find out how much N was in the leaves. A GreenSeeker monitor was used to measure NDVI, which shows how healthy and full of biomass the plants are. A SPAD meter was used to measure chlorophyll content, which is linked to nitrogen content. These readings were taken at regular times during the growing seasons to keep track of how the crops changed and improved after nitrogen was applied.

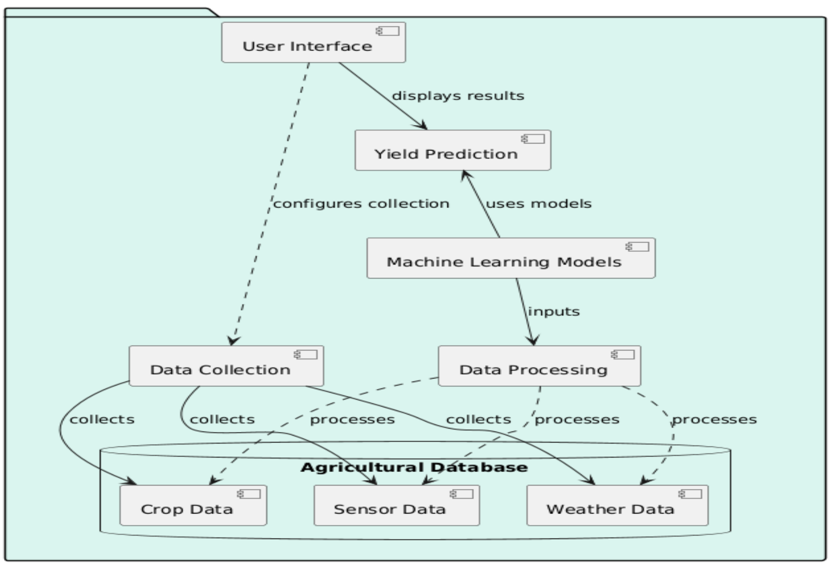

Figure 1: Overview of System architecture

The models is guided in estimating crop yields prediction by data points including NDVI, SPAD, leaf-N content, and nitrogen rates applied. By means of many predictors with model, one may assess their performance and choose the most accurate one for useful application. This method maximises nitrogen in use in food production as well as enables the forecast of crop yields at different nitrogen levels. Combining modern machine learning techniques with extensive agricultural data collection, this study offers a complete set of methods for predicting and improving crop yields under nitrogen-variable settings.

4. ALGORITHM USED

A. Support Vector Machine (SVM): Strong machine learning tool Support Vector Machine (SVM) performs well for activities like classification and regression in crop growth prediction. It operates by determining the hyperplane with the smallest gap that separates the classes in the dataset most precisely. SVM can handle variables like NDVI, SPAD, leaf nitrogen content, and crop yields that do not have a direct line link, therefore enabling Maize and Wheat nitrogen stress and yield modelling. In high-dimensional spaces it is stable and performs well when there are more dimensions than samples. This makes it ideal for handling complex agricultural data where accuracy is really crucial.

B. RF (Random Forest): Random Forest (RF) is a kind of ensemble learning in which several decision trees are used to generate more accurate predictions free from too strong fit. Combining the data from several trees helps RF to estimate more precisely and steadily. When used to forecast nitrogen stress and crop output in maize and wheat, RF can manage the great degree of variance in agricultural data and illustrate how several inputs impact each other in convoluted ways. It is particularly helpful as it allows one to rank the importance of many factors, thereby guiding the identification of the ones most influencing food development.

C. Decision Tree (DT): Decision Tree (DT) is a guided learning method that doesn't use parameters and is used for both classification and regression. You can make a tree by cutting a dataset into parts, which lets you make decisions in a structured way. A DT can make it easier to make decisions about farming yield prediction by showing clear paths from variables (like soil nitrogen levels or weather trends) to results (like crop yields). DT is helpful because it is easy to understand and use, and it gives clear information about how different factors affect results. However, it can overfit, especially when there is a lot of detailed data.

D. LR (Logistic Regression): Logistic Regression (LR) is usually used for two-class classification problems, but multinomial logistic regression and other methods can make it work for prediction problems with more than two classes. If you want to guess what crop yields will be based on nitrogen levels and other signs like NDVI and SPAD values, LR can model the chance of yield results based on these inputs. LR is simpler and less flexible than the other models we talked about, but it can be used to quickly compute odds that can help with assessing risk and making decisions in crop management.

5. RESULT AND DISCUSSION

The pairplot analysis using the given dataset clearly shows the strong positive relationships between different farming factors and nitrogen (N) application rates. This highlights how important nitrogen is for maintaining crop health and increasing yield. The data show that N is a useful mineral that raises the amount of chlorophyll, improves plant health, and eventually raises food yields.

N-Rates vs. SPAD Value: In both wheat and maize, there is a positive relationship between N-rates and SPAD values. This means that higher nitrogen supply leads to more chlorophyll. This connection is very important because chlorophyll is needed for photosynthesis, which is how plants make energy. Higher SPAD results show that as nitrogen treatment goes up, so does chlorophyll output. This link not only proves that nitrogen increases the ability of plants to make food, but it also shows that SPAD meters are useful for checking chlorophyll levels and, by extension, how healthy plants are.

N-Rates vs. NDVI: For the same reason, the link between N-rates and NDVI shows how nitrogen affects the health and vigour of plants. NDVI is a way to measure how dense and green the foliage is. It shows how much energy and health there is in the plants. The found positive relationship proves that nitrogen helps plants grow better, which results in thicker, stronger crops. In precision agriculture, this connection is especially useful for changing the amount of nitrogen used to get the best crop health and output.

N-Rates vs. Leaf N Content: It makes sense that there would be a clear link between the amount of nitrogen applied and the amount of nitrogen in the leaves. This is because plants naturally absorb and use more nitrogen when it is available. This increased nitrogen buildup helps the body do many important things, like making proteins and energy, which are needed for growth and development.

N-Rates vs. Grain and Straw Yield: The fact that grain and straw yield went up as N application rates went up shows that nitrogen helps photosynthesis and makes it easier for plants to move and store carbohydrates. This finding is important because it shows that managing nitrogen correctly can have a direct effect on crop yield, which is necessary to meet the needs of food production.

SPAD vs. NDVI, SPAD vs. N Content, and SPAD vs. Straw Yield: The relationships between SPAD, NDVI, leaf N content, and straw yield show how these factors are linked. Higher NDVI and leaf N content are linked to higher chlorophyll content (SPAD). These levels show that the plant is healthy and growing strongly. Because of this, these factors lead to higher biomass and yield, showing that nitrogen has a broad effect on crop growth from both a physiological and an agricultural point of view.

Table 2: Performance Metrics of Machine Learning Models for Crop Yield Prediction Based on Nitrogen Management

Model	MAE (kg/ha)	MSE (kg/ha²)	RMSE (kg/ha)	R^2
SVM	45.2	3089.1	55.6	0.92
RF	35.0	2100.5	45.8	0.95
DT	50.3	4025.4	63.4	0.89
LR	60.1	4800.7	69.3	0.86

Table 2 shows how well Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR) four machine learning models work at predicting crop yields based on nitrogen management. The Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) were used to test these models. These are important ways to see how accurate and useful forecast models are in agriculture. With a R^2 of 0.92, an MAE of 45.2 kg/ha, an MSE of 3089.1 kg/ha², and an RMSE of 55.6 kg/ha, the SVM model works well, as illustrate in figure 2. These measures show that SVM works very well, making accurate predictions about yields with a high level of agreement between actual and projected values.

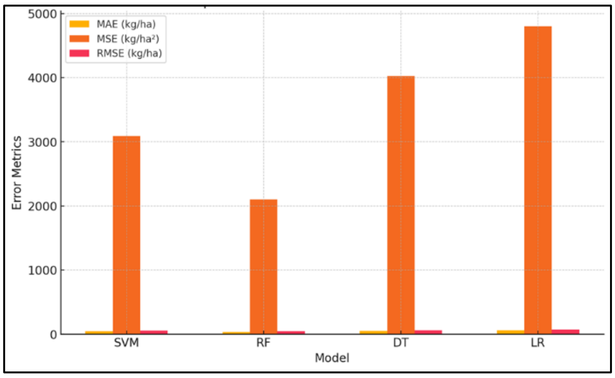

Figure 2: Representation of Evaluation metrics comparison for ML Models

The low mistake rates and high R^2 show that it does a good job of handling the different types of data, which makes it a good choice for this task. With an MAE of 35.0 kg/ha, an MSE of 2100.5 kg/ha², an RMSE of 45.8 kg/ha, and a R^2 of 0.95, the RF model does better than the others. RF's better success is due to its ensemble method, which averages several decision trees learnt on different parts of the same dataset to lower error and bias. This makes RF a great choice for farming datasets that are hard to understand because the relationships between factors can be complicated and not linear.

The DT model, on the other hand, has a smaller R^2 of 0.89 and higher error rates (MAE of 50.3 kg/ha, MSE of 4025.4 kg/ha², and RMSE of 63.4 kg/ha). It is also simpler and easier to understand. Overfitting can cause the mistakes to be higher because the model learns noise from the training data instead of the real links. The LR model has the worst scores of the four. Its MAE is 60.1 kg/ha, its MSE is 4800.7 kg/ha², its RMSE is 69.3 kg/ha, and its R^2 is 0.86. This is despite the fact that it is usually simpler and faster to use. LR may be fast at computing, but it has trouble with the complicated information used for nitrogen management and crop yield forecasts because it has connections that don't follow a straight line.

Table 3: Comparison of accuracy and precision metrics for various machine learning models applied in crop yield prediction

Model	Accuracy (%)	Precision (%)
SVM	92.5	91.0
RF	95.8	94.2
DT	88.7	87.5
LR	85.4	84.9

Table 3 shows a quick summary of the precision and accuracy measures for four machine learning models used to guess crop yields: Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR). These measures are very important for judging how well forecasting models work, especially in precision agriculture where getting the most out of inputs and inputs is very important. With a 92.5% accuracy rate and a 91.0% precision rate, the SVM model does a great job.

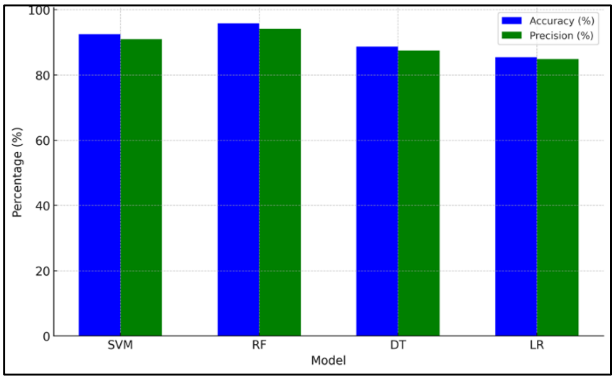

Figure 3: Comparison of Accuracy and Precision

Based on these numbers, it looks like SVM is pretty good at correctly classifying the yield results based on the inputs and doing this consistently across different data points. Its high accuracy means that it doesn't give many fake results, which is important to keep food amounts from being overestimated. Random Forest has the best accuracy and precision in the group, with scores of 95.8% for accuracy and 94.2% for precision. This better performance shows that RF can handle complicated datasets with variables that are linked to each other. This is common in farming data because of the many natural factors that affect crops. It's clear that RF's ensemble method, which uses multiple decision trees to prevent overfitting and improve the model's ability to be used in different situations, works because it makes estimates that are more accurate and reliable. Even though Decision Tree is easier to understand and use, it is not as accurate or precise as SVM and RF (88.7% vs. 87.5%, respectively). This could be because it's easy for it to become too good at what it does, especially when working with complicated farming records where a lot of things affect the result. With an accuracy of 85.4% and a precision of 84.9%, Logistic Regression seems to be the least useful of the four models. This could be because LR is linear and might not be good at dealing with the non-linear interactions that are common in farming datasets.

6. CONCLUSION

The work has effectively shown the great possibilities for optimising nitrogen management and raising crop yields by combining machine learning methods with agricultural practices. The work used machine learning models including SVM, Random Forest, Decision Tree, and Logistic Regression to forecast the yields of maize and wheat using a dataset enhanced with factors like leaf nitrogen content, NDVI, SPAD values, and varied nitrogen application rates. With the greatest accuracy and precision, Random Forest proved better than other models, therefore demonstrating its capacity to manage difficult data and derive important insights from agricultural data. This emphasises the resilience of the model and its fit for precision farming, in which accurate decision-making is very vital. Another dependable approach for estimating crop outcomes is SVM, which likewise shown good performance especially in managing nonlinear connections between the variables. Since both the quantity and timing of nitrogen given to crops greatly affect yield results, this study emphasises the need of accuracy in nitrogen administration. Direct measurements of leaf nitrogen content, NDVI, and nitrogen rates as well as their favourable correlations with SPAD values establish the efficacy of these indices as markers of crop health and nitrogen sensitivity. Using predictive analytics in agriculture presents a revolutionary way to control crop nutrition, maximise resource utilisation, and raise agricultural output. Stakeholders can better address the dual issues of increasing agricultural yields and preserving environmental health by keeping improving these models and combining them with real-time data collecting tools.

REFERENCES

[1] Friha, O.; Ferrag, M.; Shu, L.; Maglaras, L.; Wang, X. Internet of Things for the Future of Smart Agriculture: A Comprehensive Survey of Emerging Technologies. IEEE/CAA J. Autom. Sin. 2021, 8, 718–752.
[2] Subeesh, A.; Mehta, C. Automation and digitization of agriculture using artificial intelligence and internet of things. Artif. Intell. Agric. 2021, 5, 278–291.
[3] Lakshmi, G.; Asha, P.; Sandhya, G.; Sharma, S.; Shilpashree, S.; Subramanya, S. An intelligent IOT sensor coupled precision irrigation model for agriculture. Meas. Sens. 2023, 25, 100608.

[4] Foughali, K.; Fathallah, K.; Frihida, A. Using Cloud IOT for disease prevention in precision agriculture. Procedia Comput. Sci. 2018, 130, 575–582.

[5] Adamchuk, V.; Hummel, J.; Morgan, M.; Upadhyaya, S. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 2004, 44, 71–91.

[6] Shafi, U.; Mumtaz, R.; García-Nieto, J.; Hassan, S.; Zaidi, S.; Iqbal, N. Precision agriculture techniques and practices: From considerations to applications. Sensors 2019, 19, 3796.

[7] Singh, M.; Karada, M.S.; Rai, R.K.; Pratap, D.; Agnihotri, D.; Singh, A.K.; Singh, B.K. A Review on Remote Sensing as a Tool for Irrigation Monitoring and Management. Int. J. Enviorn. Clim. Chang. 2023, 13, 203–211.

[8] Liu, Y.; Zhang, X.; Xi, L.; Liao, Y.; Han, J. Ridge-furrow planting promotes wheat grain yield and water productivity in the irrigated sub-humid region of China. Agric. Water Manag. 2020, 231, 105935.

[9] Wei, T.; Dong, Z.Y.; Zhang, C.; Ali, S.; Chen, X.L.; Han, Q.F.; Zhang, F.C.; Jia, Z.K.; Zhang, P.; Ren, X.L. Effects of rainwater harvesting planting combined with deficiency irrigation on soil water use efficiency and winter wheat (Triticum aestivum L.) yield in a semiarid area. Field Crops Res. 2018, 218, 231–242.

[10] Zhang, P.; Wei, T.; Han, Q.F.; Ren, X.L.; Jia, Z.K. Effects of different film mulching methods on soil water productivity and maize yield in a semiarid area of China. Agric. Water Manag. 2020, 241, 106382.

[11] El-Hendawy, S.E.; Al-Suhaibani, N.A.; Elsayed, S.; Hassan, W.M.; Dewir, Y.H.; Refay, Y.; Abdella, K.A. Potential of the Existing and Novel Spectral Reflectance Indices for Estimating the Leaf Water Status and Grain Yield of Spring Wheat Exposed to Different Irrigation Rates. Agric. Water Manag. 2019, 217, 356–373.

[12] Eugenio, F.C.; Grohs, M.; Venancio, L.P.; Schuh, M.; Bottega, E.L.; Ruoso, R.; Schons, C.; Mallmann, C.L.; Badin, T.L.; Fernandes, P. Estimation of Soybean Yield from Machine Learning Techniques and Multispectral RPAS Imagery. Remote Sens. Appl. 2020, 20, 100397.

[13] Ramos, A.P.M.; Osco, L.P.; Furuya, D.E.G.; Gonçalves, W.N.; Santana, D.C.; Teodoro, L.P.R.; da Silva Junior, C.A.; Capristo-Silva, G.F.; Li, J.; Baio, F.H.R.; et al. A Random Forest Ranking Approach to Predict Yield in Maize with Uav-Based Vegetation Spectral Indices. Comput. Electron. Agric. 2020, 178, 105791.