International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Artificial Intelligence-Enabled Sustainable Crop Growth Optimization via Comprehensive Environmental Data Analysis

Dr. B. Bhanu Prakash[1], P. Vijay Ganesh[2], J. Naga Krishna[3], N. Chiranjeevi[4], S. Sai Tarun[5], Sk. Riyaz[6]
[1]Professor & Head of Department, Dept. of CSE–Data Science
KKR & KSR Institute of Technology and Sciences, Guntur, India
Email: prakashbattula33@gmail.com [1]
[2,3,4,5,6]B.Tech Students, Dept. of CSE–Data Science
KKR & KSR Institute of Technology and Sciences, Guntur, India
Emails: pavulurivijayganesh@gmail.com[2], nagakrishna0822@gmail.com[3], 23jr5a4413@gmail.com[4], 22jr1a44b9@gmail.com[5], 22jr1a44c5@gmail.com[6]

Abstract—

Agriculture is one of the main pillars of food pro- duction in the world, which is experiencing increasing problems because of the variability of climatic conditions, poor use of available resources and decreasing soil fertility. Conventional farming practices are prone to manual surveillance and judgment which are inaccurate, labour-intense practices incapable of keep- ing up with the dynamic environmental factors. Crop monitoring and yield prediction have been implemented using the existing methods like rule-based systems and classical machine learning like Decision Trees and Random Forests. These models however do not usually account for complex nonlinear relationships and temporal dynamics of the environmental and soil data to make optimum decisions and to be scalable. In order to address these shortcomings, the current research suggests sustainable crop growth optimization system based on Artificial Intelligence and a CNN BiLSTM deep learning model. Convolutional Neural Networks (CNN) are used in the model to obtain spatial relation- ships between features, including soil moisture, pH, temperature, humidity, and light intensity, and the Bidirectional Long Short- Term Memory (BiLSTM) element is used to obtain forward and backward temporal relations among sensor data sequences. The dataset, which will be used, is Smart Agriculture and Plant Health Monitoring using IoT, which offers multivari- ate environmental measurements. Experimentally, it is shown that CNN-BiLSTM model works better in prediction accuracy, temporal stability and generalization to achieve considerably higher improvements in root mean square error (RMSE) and mean absolute error (MAE). The suggested model is effective in predicting the best irrigation and environment changes to ensure the sustainable management of the resources, reduction of waste of water and other fertilizer, and increase of crop production and ecological stability. Such a strategy opens the path to smart precision farming and data-driven sustainable farming.

***Index Terms*—Smart Agriculture, CNN–BiLSTM, Deep Learn- ing, Environmental Data Analysis, Sustainable Farming**

I. Introduction

Agriculture has been the foundation of food security in the world but is under constant pressure because of the variability of climatic conditions, declining soil quality, water scarcity and the necessity of sustainable productivity. Manual monitor- ing and decision-making allowed in the past has made farmers reliant on empirical judgment which in most cases results in inefficient irrigation, unreliable crop production and over in- corporation of fertilizers. As the need to become more precise and sustainable, the application of Artificial Intelligence (AI) to the agricultural field provides a revolutionary opportunity since it allows using the analysis of the environment and the state of the soil to manage crops optimally.

Although the current computational agriculture has some improvement, there are a number of limitations with its current systems. Conventional methods like decision systems that are rule based and statistical regression model do not represent the multi-factor interactions, which are intricate and nonlinear involving several environmental conditions such as soil pH, humidity and temperature. Classical machine learning algorithms such as Decision Trees, Support Vector machines (SVM) and Randon Forests have been extensively applied in crop prediction and yield analysis. Nevertheless, these models are very reliant on feature engineering and are unable to model temporal dependencies between environmental data streams. Though the deep learning models including LSTM and CNNLSTM hybrids have demonstrated potential in time series pattern learning, they tend to be susceptible to problems like gradient vanishing, overfitting and loss of contextual knowledge in small or noisy IoT signals.

II. Problem Statement

To address these limitations, the study will offer a Convolu- tional Neural Network-Bidirectional Long Short-Term Mem- ory (CNN-BiLSTM) hybrid model to ensure optimal growth of crops in a sustainable manner. The proposed model uses CNN layers to draw spatial correlations among features of temperature, soil moisture, and light intensity whereas the BiLSTM layers find bidirectional time-series sensor measure- ments. The Smart Agriculture and Plant Health Monitoring using IoT dataset is used to train the model and it includes a variety of environmental parameters that are important to provide an accurate decision support in crop management.

CNNBiLSTM architecture has shown to have higher predictive accuracy, robustness and flexibility compared to classical and unidirectional deep learning models thus enhancing irrigation scheduling, fertilizer optimization and real time environmental monitoring.

III. Objectives

- To create and train a CNN-BiLSTM deep learning network to optimize the growth of crops with the help of multivariate environmental data.
- To examine and process sensor environmental data to determine key parameters that influence crop production.
- To make the process more sustainable, by reducing water and fertilizer wastage with accurate prediction and recommendation through AI.
- To compare the working of the proposed model with the current ML and DL arch models in terms of accuracy, RMSE, and MAE.
- To emerge with the scalable framework that can respond to the dissimilar crops, geographical regions, and agriculture environments.

The rest of the paper will be structured in the following way: Section 2 will be a literature review Section 3 will outline the dataset and the methodology that will be used in the experiment, Section 4 will discuss the experimental findings and conclusion, and Section 5 will be a conclusion of the paper with prospective directions.

IV. Literature Survey

The application of machine learning (ML) and deep learning (DL) technologies in agriculture has turned into one of the enablers of smart farming and sustainable crop management. The use of environmental, soil, climatic and IoT sensor data has allowed researchers to produce smart systems that can predict the crop growth, manage the irrigation process and make data-driven decision-making. Although these have improved, there are still issues of limited temporal modelling, poor extrapolation to unknown states and high computational cost that still limit large-scale application to real-world problems and scenarios.

Talaat [1] suggested Crop Yield Prediction Algorithm (CYPA) that combines the data of the IoT sensors with the climate information to ensure accuracy in agriculture. This system showed better irrigation planning and yield estimation, although it depended on the traditional machine learning models, which could not represent the complex nonlinear and long-term temporal relations among environmental parameters. On the same note, Abdel-Salam et al. [2] came up with a hybrid model that integrates feature selection with optimized machine learning methods to improve prediction accuracy. Although effective, the method involved a lot of feature engineering, and was not flexible to environment dynamically changing conditions.

In order to deal with these drawbacks, deep learning methods are progressively being studied. El-Kenawy et al. [3] used machine learning and deep learning models to predict the yield of potato crops and found that these methods performed better than traditional ML methods. The study however indicated difficulties in extrapolating the model between climatic regions. Boukhris et al. [4] employed satellite imagery, IoT data and deep learning to predict the yield of wheat in Morocco. Even though the method was very accurate spatially, it was a resource-intensive computation and required large-scale datasets, thus hampering its applicability to resource-constrained agricultural settings.

Neural learning has also been studied based on optimization to enhance its prediction performance. The authors of Jovanovic et al. [5] compared metaheuristic-tuned weight-agnostic neural networks with crop yield prediction and found a higher degree of accuracy and robustness. The higher computational cost, however, became a challenge to the real-time IoT-based agricultural application. Subramaniam and Marimuthu [6] suggested a deep learning model with dimensionality reduction of Indian regional crops, which increased accuracy and minimized redundancy of features. The model did not, however, explicitly model the bidirectional temporal dependencies, and was largely based on a situation in which feature representations were static.

Machine learning has also been used in regional crop modeling. In [7], Seck et al. used localized modeling because of the application of environmental data to predict crop yield in Senegal using AI-based methods. Nevertheless, the model did not allow it to capture the long-term environmental trends because of the lack of deep temporal learning. Sharafat et al. [8] designed an AI system based on the IoT system to predict crops using soil and weather measurements in real-time. Although efficient in the short-term prediction and monitoring, the unidirectional temporal modelling was limiting the ability to completely utilise the past and the future environmental dependencies.

V. Research Gaps

The research gaps identified based on the study are:

1) The majority of current ML and DL models do not effectively address the bidirectional time-based relationships in the environmental and sensor-based agricultural data.
2) The complexity of advanced deep learning models fails to be implemented on low-power IoT systems due to its high computational complexity.
3) Most techniques are not applicable in different climatic and soil conditions.
4) There is often a lot of hyperparameter tuning and preprocessing in hybrid models, and this makes them more difficult to implement.
5) There is a low focus on sustainability-concerned results, like optimised use of water and fertilisers.

These constraints drive the creation of CNN-BiLSTM hybrid that can jointly learn spatial correlations and bi-directional temporal connections between multivariate environmental data to allow correct, scalable and sustainable optimization of crop growth.

VI. Proposed System

This part shows the procedure of the creation of the CNN–BiLSTM-based optimization system of sustainable crop growth. It entails data collection, pre-processing and development of a deep learning model comprising both spatial and temporal analysis of environmental data. The given framework is supposed to make farming more sustainable, combining the information provided by environmental sensors, deep learning algorithms, and decision support analytics to optimise the irrigation options, fertiliser usage, and forecasts of the crop productivity.

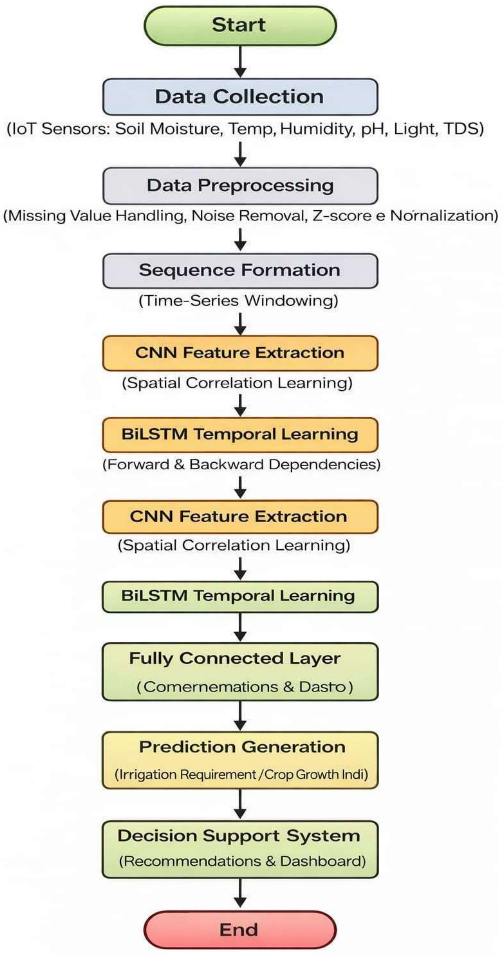

Fig. 1. Overall architecture of the proposed CNN–BiLSTM-based system for sustainable crop growth optimization using IoT environmental data.

The general structure of the suggested CNN-BiLSTM model of sustainable crop growth optimization. The framework starts with the real time data collection of the environmental IoT sensors, which measure soil moisture, temperature, humidity, rainfall, light intensity as well as soil pH. Pre-processing of these readings is then done to eliminate noise, deal with missing data and normalize all the features to a homogenous scale using Z-score standardization where data of the same nature is of similar quality. The data is sent to CNN-BiLSTM hybrid network after the preprocessing step. The CNN layers are used to obtain local spatial correlations among environmental variables (such as the corrosion of temperature and humidity on soil moisture), and the BiLSTM layers are used to obtain bidirectional temporal dependencies, or what past trends of the environment users can tell about the future. The combination of the two models enables spatial and time learning to enable sound and dynamic crop management. Lastly, the dense layer generates some predictive results like the best irrigation rates, nutrient prescriptions, or the anticipated crop growth rates. These outputs are presented in an easy-to-use dashboard to enable farmers to make data driven decisions. The architecture incorporates both environment monitoring and time prediction and responsive analytics to enhance sustainability, productivity and resilience of crop production.

A. Data Acquisition

The article uses the publicly available Smart Agriculture and Plant Health Monitoring using IoT dataset containing constant multivariate environmental measurements measured through real-life sensors. The data set consists of the following parameters:

- Soil Moisture (%): Reflects water content which is important in schedule of irrigation.
- Soil Temperature (°C): Influences absorption of nutrients and microorganisms.
- Ambient Temperature (°C) and Humidity (%): Indicates weather patterns that affect the process of evapotranspiration.
- Light Intensity (Lux): Light Intensity is the measure that determines the photosynthetic efficiency.
- Soil pH: Regulates nutrient content and health of crops.
- Water TDS (ppm): Refers to the dissolved minerals in the soil which influence the soil fertility.
- Battery Voltage: This is to ensure the sensor remains stable to do continuous monitoring.

The individual sensors produce time-series signals, which are sampled at a regular time scale, which allows crop-environment interactions to be studied over a long term.

B. Preprocessing and Augmentation of Data

In order to achieve model accuracy and generalization, several steps of data preprocessing have been carried out:

1) Missing Value Imputation:: Sensor readings have sometimes gapped in them because of transmission or calibration problems. Linear interpolation and temporal smoothing were used to fill in the missing data to provide continuity to the time-series data.

2) Normalization:: Scaled each feature using Z-score normalization so that they would contribute equally in training:

$$z = \frac{x - \mu}{\sigma}$$

Where x= feature value, μ = mean, and σ is standard deviation of the feature.

3) Sequence Formation:: The data was split in 24-hour rolling time windows, and the model was trained to make a prediction of the next hours or the next day of the environmental variable (soil moisture or growth index).

4) Data Augmentation:: Gaussian noises have been added to the training data to ensure there was a slight perturbation that enhanced generalization and avoiding overfitting. The model was modified with the help of the augmented dataset so that it adjusted itself to natural environmental fluctuations.

C. CNN–BiLSTM Framework

The CNN-BiLSTM hybrid is structured in a manner that it utilizes space and time-based features of environmental measurements to optimize sustainable crop development.

- Convolutional Layers (CNN): Extrapolate spatial relationship and feature patterns of correlated parameters of temperature, humidity and soil moisture. The convolutional filters automatically capture higher-level representations that are of crop health.
- Bidirectional LSTM (BiLSTM): Temporal dependencies are calculated both forward and backward, as it tends to learn long-term and short-term changes in the environmental factors.
- Dropout Layers: Added to decrease the overfitting and enhance the generalization.
- Fully Connected Layer: Transforms learned representations into particulars like yield forecasts, irrigation requirements, or crop strain signals.

Mathematically, the BiLSTM updates can be represented to be:

$$h_t = \text{BiLSTM}(x_t, h_{t-1})$$

$$y_t = \sigma(Wh_t + b)$$

with input sequence x_t denoted by x_t, hidden state by h_t, weight matrix denoted by W, bias denoted by b and activation function denoted by σ.

VII. Methodology

1) Gather live farm information on IoT sensors.
2) Gather multivariate sensor data like the soil moisture, soil temperature, soil humidity, water light intensity, soil pH and water TDS levels.
3) Preprocess the data gathered to eliminate noise, process missing data and normalize features.
4) Sliding window techniques used to form time-series sequences.
5) Infer relational features of spatial features with Convolutional Neural Networks (CNN).
6) Learn bi-directional temporal dependencies with BiLSTM.
7) Make forecasts of irrigation demand and the indicators of crop growth.
8) Give decision support advice to sustainable farming.

Predicting crop growth in a specified area with the help of a CNN-BiLSTM model [1] Input: Data on IoT sensors (soil moisture, temperature, humidity, light intensity, pH, TDS) Output: Forecasted irrigation demand / crop development marker. Read IoT multivariate sensor data. Deal with missing values, eliminate noise in sensor values. Normalize sensor data with Z-score normalization. Make time-series sequences during sliding window strategy. Taking up spatial features with CNN layers. Temporal patterns Learn BiLSTM layers. Make prediction with fully connected layer. Predicted output of crops or quantity of irrigation needed.

Data pre-processing and Decision Support. [1] Input: Raw data of environmental sensors. Output: Recommendations of actionable crop management. Get raw sensor IoT data. Carry out noise removal and missing value imputation. The scaling of all features should be normalised. Organize data after time in sequence. Get prediction outputs of CNN-BiLSTM model. Compare the values of prediction against optimal threshold values. Produce irrigation and crop control prescriptions. Present the recommendations using decision-support dashboard.

VIII. Results and Discussion

The CNN-BiLSTM system optimization system of sustainable crop growth was tested on the Smart Agriculture and Plant Health Monitoring using IoT dataset. As the model is developed as a regression problem, its performance has been observed in terms of the best prediction of continuous environmental and crop-related variables, e.g. soil moisture level, irrigation needs, etc. The performance of the model was evaluated through the comparison of its performance with the traditional machine learning and deep learning methods.

A. Performance Metrics

To measure the performance of the proposed regression model, two popular error-based measures were used, which include Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics are the means of the comparison between the values that the model produces and the real ground truth values.

- Mean Absolute Error (MAE): This denotes the mean of the absolute error between the actual and the predicted values. It gives a clear signal of the proximity of the predictions to the exact values without putting any additional emphasis on the contribution of large errors.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

- Root Mean Square Error (RMSE): This is a square root of average of the squared errors between the predicted and actual values. RMSE perceives larger errors more severely than MAE and hence it is applicable in assessing model stability.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

where y_i represents the actual value and $\hat{y}_i$ denotes the predicted value.

B. Experimental Results

Model	MAE	RMSE
Decision Tree	0.92	1.18
Random Forest	0.71	0.94
LSTM	0.56	0.73
CNN-LSTM	0.44	0.61
CNN-BiLSTM	0.32	0.48

Fig. 2. Performance comparison of the proposed CNN-BiLSTM model with baseline models.

Table 1 shows the performance of the proposed CNN-BiLSTM model compared to the baseline machine learning and deep learning models.

The findings clearly point out that the traditional machine learning models that include Decision Tree and Random Forest yield more prediction errors because of the low predictive power of the model to develop temporal relationships that exist within environmental data. The LSTM model works better by learning time sequences; it also analyses information in a forward time manner. The CNNLSTM model also increases the accuracy of predictions by adding the features of space extraction, yet it does not have full temporal elements. Comparatively, CNN-BiLSTM model has the smallest values of MAE and RMSE, which is more accurate and stable in predicting continuous environmental variables.

C. Loss vs Epochs Analysis

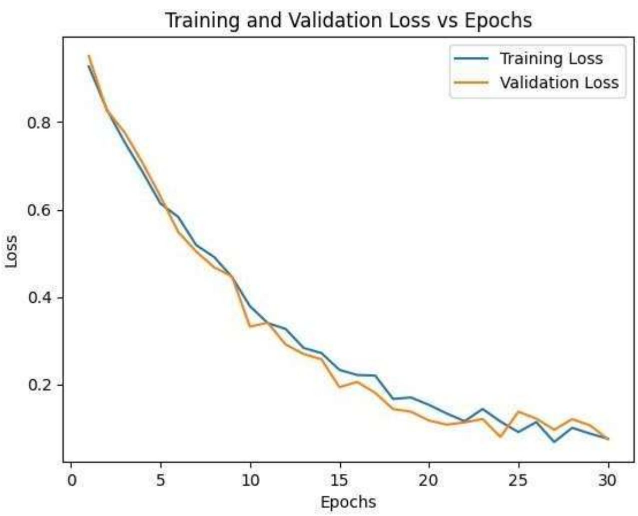

Fig. 3. Training and validation loss curves for the CNN-BiLSTM model.

Fig. 3 shows how training and validation loss change as the number of training epochs of CNN-BiLSTM model increases. The initial epochs have higher values of loss as the features are not learned well. The loss is gradually reduced as the model acquires spatial-temporal representations. Training and validation loss curves meet, which implies the stability of learning. The lack of divergence establishes the lack of overfitting. This discussion shows that the suggested model is convergent and can be easily generalized to unknown data.

D. Comparison with the Existing Methods of Performance

Method	**Prediction Accuracy**	**Automation**	**Computational Cost**	**Real-Time Capability**
Manual Monitoring	Low	No	Low	No
Rule-Based Systems	Medium	Partial	Medium	Limited
ML-Based Models	Medium	Yes	Medium	Yes
Proposed CNN - BiLSTM	High	Yes	Medium	Yes

Fig. 4. Performance comparison with existing approaches.

Table 2: Comparison with Existing Approaches.

The suggested CNN-BiLSTM model is more effective than the traditional methods which offer better prediction accuracy, complete automation and real-time decision support with a reasonable computational cost.

E. Discussion

The enhancement in the CNN-BiLSTM model can be explained by the fact that the hybrid architecture is able to capture the spatial relationships, as well as the bidirectional temporal relationships in the multivariate environmental data. The CNN layers automatically acquire interrelationships between correlated parameters like temperature, humidity, and soil moisture and the BiLSTM layers capture all the long-term temporal relationships in both the past and future. The bidirectional temporal modeling works much better than unidirectional LSTM-based techniques in explaining the patterns of seasons as well as abrupt changes in the environment. This leads to a smaller error in prediction and better stability, especially in the case of noisy data of the IoT sensors.

Sustainably, proper regression-based forecasting of the irrigation needs will allow the optimal use of water and fertilizers. The proposed system helps to reduce resources wastage and encourages farming practices that are ecologically friendly and enhance the soil health in the long run. Moreover, the trade-off between accuracy of predictions and computational efficiency is balanced, which makes the proposed model applicable to be deployed to the IoT-enabled precision agriculture systems.

IX. Expected Outcomes

- Precise forecasting of constant crop growth indicators.
- Smart irrigation need predictions.
- Minimized wastages of water and fertilizers.
- There is better stability of crop yield.
- Less reliance on manual decision making.
- Scalable and Sustainable Smart Agricultural Solution.

X. Conclusion

This study described a CNN-BiLSTM hybrid deep learning model with the implementation of Artificial Intelligence-based optimization of sustainable crop growth. Combining the convolutional neural networks of spatial features with the bidirectional LSTM networks of time-dependent features, the proposed system analysed complex interactions of the environment and time-dependent patterns of multivariate IoT sensor data more effectively, represented by lower values of MAE and RMSE. Experimental analysis conducted on the Smart Agriculture and Plant Health Monitoring using IoT dataset showed that the CNN-BiLSTM model achieved higher accuracy, stability, and generalization compared to traditional machine learning and unidirectional deep learning models. The enhanced predictive potential will allow effectively planning irrigation and environmental responses and ultimately reducing water and fertilizer waste, increasing crop production, and improving the sustainability of the soil. The proposed framework will promote sustainable farming by enabling effective data-driven decision-making, enhancing the efficiency of resource use, and creating more robust ecological conditions. Its scalable and flexible architecture makes it compatible to be deployed in various crops, climatic conditions and IoT-enabled farming-based systems.

Further research will be conducted to expand the framework to include attention mechanisms and transformers models to improve temporal learning. Also, it can be extended by the edge computing deployment of real-time inferences and the investigation of the possibilities of reinforcement learning-based optimization strategies that will enhance the results of system autonomy and sustainability.

REFERENCES

[1] Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Computing and Applications, 35(23), 17281-17292.

[2] Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. Neural Computing and Applications, 36(33), 20723-20750.

[3] El-Kenawy, E. S. M., Alhussan, A. A., Khodadadi, N., Mirjalili, S., & Eid, M. M. (2025). Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. Potato Research, 68(1), 759-792.

[4] Boukhris, A., Jilali, A., & Sadiq, A. (2024). Satellite imagery, big data, IoT and deep learning techniques for wheat yield prediction in Morocco. Results in Control and Optimization, 17, 100489.

[5] Jovanovic, L., Zivkovic, M., Bacanin, N., Dobrojevic, M., Simic, V., Sadasivuni, K. K., & Tirkolaee, E. B. (2024). Evaluating the performance of metaheuristic-tuned weight agnostic neural networks for crop yield prediction. Neural Computing and Applications, 36(24), 14727-14756.

[6] Subramaniam, L. K., & Marimuthu, R. (2024). Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 8, 100611.

[7] Seck, N. K. G., Ngom, A., Ngom, P., & Noba, K. (2025). Crop yield forecasting in Senegal: application of machine learning methods. Discover Agriculture, 3(1), 1-18.

[8] Sharafat, M. S., Kabya, N. D., Emu, R. I., Ahmed, M. U., Onik, J. C., Islam, M. A., & Khan, R. (2025). An IoT-enabled AI system for real-time crop prediction using soil and weather data in precision agriculture. Smart Agricultural Technology, 101263.

[9] Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science, 10, 1750.

[10] Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldu´, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and electronics in agriculture, 143, 23-37.

