Towards Smart Farming: An Implementation-Oriented Approach Leveraging Data-Driven Agricultural Systems

Siya Sharma

IMS, Noida

sharmasiya532@gmail.com

ABSTRACT: **Although the agricultural sector is fundamental to global food security, agri-food systems are often perceived as providing low, wasteful, and unpredictable societal value. Agriculture using data driven technologies rooted in the Internet of Things (IoT) and machine-learning provide a new and hopeful mechanism for improving productivity and profitability, and by extension, sustainability of agricultural systems. The referenced article features an implementation-oriented framework for using IoT wired sensors, cloud storage, and model predictive for smart agriculture applications. Low-cost wired sensors were used to gather in real-time soil, meteorological, and crop health data, which were then uploaded to a cloud for preprocessing and analytics. Twelve separate machine leaning models were utilized, including Random Forest (RF) for irrigation prediction, regression analyses for yield, and convolutional neural networks (CNN's) for disease prediction, which provided action insights in time. Experimental data showed 92% irrigation prediction accuracy, 8% mean absolute error for yield estimation, and disease prediction classification accuracy of 95%. These data support the idea that data driven agriculture can facilitate decision making, improve resource use efficiency and contribute to sustainable agriculture.**

Keywords: *Data-driven agriculture, IoT, Machine learning, Precision farming, Crop yield prediction, Smart farming, Disease detection, Cloud computing*

I.INTRODUCTION

Historically, agriculture has depended on human experiences, knowledge that was specific to the season, and physical observations. Due to rising food needs, changing climates, and resources limitations, it is clear that traditional farming is becoming progressively less helpful [1]. Therefore, farmers are frequently relying upon guess work on day-to-day decisions such as irrigation cycles, fertilizers, and pest management. Guessing implies that the farmer is likely to apply extra inputs that are not necessary which results in reduced yields and profitability [2].

Data-driven agriculture is a new opportunity to convert agriculture into a principle-based, efficient, precise, and sustainable practice [3]. Data-driven agriculture provides an opportunity to collect agricultural data in a systematic way from multiple sources, including IoT sensors, satellite-derived data, and weather stations, which can be processed, analyzed, and modelled to aid in deciding on best practices for farm management. Unlike previous approaches, data-driven systems take advantage of machine learning algorithms and computing capabilities offered through the cloud to allow farm management practices to integrate "real-time" actionable insights [4][5][8].

Through the use of data-driven agriculture, farmers are able to:

- Assess field-level soil and crop health in real-time.
- Enhance fertilizer and water use efficiencies through predictive analytics.
- Identify diseases and pest pressures in crop fields at early stages.
- Provide productivity estimates to assist with marketing forecasts.

This paper discusses an implementation-based study of data-driven agriculture using IoT-enabled data acquisition, cloud-based data storage, and machine learning models used for predictive analytics and decision support [9]. The objective of this data-driven system is a low-cost sensor interfacing with microcontrollers for data acquisition and subsequent relay to a cloud-based platform, which will store and process data toward insights for irrigation, fertilization, and crop disease identification [9][10].

In this study, we propose that a data-driven farming system can be implemented in a practical and operational farming environment. The results reported in this paper provide evidence of the feasibility of data-driven farming systems to improve productivity, reduce input costs, and support the sustainability of agricultural practices [3][7].

II.SYSTEM ARCHITECTURE

2.1 Network and Processing Layer

This layer operates as the middleware between sensors and decision-making modules; data transmission, storage, and preprocessing occur in this layer.

- Data Transmission: IoT devices transmit data packets to the cloud via Wi-Fi/LoRa gateways.
- Cloud Storage: Data is stored on Firebase/AWS IoT Core.[11][12]
- Data Preprocessing: In this layer, preprocessing noise removal, normalization, and feature engineering are accomplished using Python (Pandas, NumPy).

Machine Learning Models:

- Random Forest for irrigation prediction.
- Regression/XGBoost for yield prediction.
- CNN for crop disease detection. [13][14]

2.2 Application Layer (Decision Support Layer)

The real-time insights and recommendations will be provided to farmers through a dashboard or mobile app.

- Visualization: Farmers will have a dashboard, e.g. with Grafana or Power BI, that allows them to monitor soil and weather parameters.
- Decision Support: Predictions made by the ML model will be converted to actionable recommendations. [15][16]

Farmer Alerts: Alerts (via SMS/Mobile App) to farmers, for example:

- “Soil moisture is low, irrigate in the next 6 hours.”
- “Disease detected: Apply fungicide immediately.” Estimated Wheat yield is 2.3 tons/acr.

III.METHODOLOGY

The methodical structure in this study uses a four-phase implementation strategy: data acquisition, preprocessing, predictive modelling, and decision support. An overview of the overall workflow is presented in Figure X (System Workflow).

3.1 Phase 1: Data Acquisition

A crucial part of the study focused on collecting real-time data from the agricultural field using IoT-enabled sensors and cameras in Phase 1. The following parameters have been measured in this phase:

- Soil Parameters: soil moisture, soil temperature, and soil pH.
- Climatic Conditions: air temperature, relative humidity, and total rainfall.
- Crop Health: images of plant leaves taken through the camera module to detect diseases.

Hardware Used:

- ESP32/Arduino was used as the microcontroller to integrate the sensors.
- DHT22 sensor for temperature and humidity.
- Soil moisture and pH sensors for soil properties.
- Camera module (Raspberry Pi Camera or ESP32-CAM) to take images of the leaves.
- LoRa/Wi-Fi modules for data transmission.

The data was transmitted to the cloud platform (Firebase/AWS IoT Core) every 10 minutes interval.

3.2 Phase 2: Data Preprocessing

The data we collected in the earlier phases was usually messy with missing values, noise, or erroneous outliers. Our preprocessing steps consisted of the following:

- Data cleaning step: Duplication and erroneous readings were removed.
- Interpolation: Missing sensor values were filled in.
- Normalization: Features had to have a similar range (0-1), therefore scaling was performed.

Feature engineering:

- Soil moisture index = (Moisture reading / Max moisture capacity).[17][18]
- Growing Degree Days (GDD) = ∑ (Daily mean temp – base temp).
- For vegetation indices (NDVI), we based the computation on the leaf images.

3.3 Phase 3: Predictive Modelling

Machine learning was used for a range of agricultural-related decision-making tasks described below:

Irrigation Prediction.

- Input: Soil moisture, temperature, humidity, rainfall.
- Algorithm: Random Forest Classifier.
- Output: Binary Decision whether irrigation is needed (Yes/No).

Crop Yield Prediction.

- Input: Soil parameters, weather, crop growth stage.
- Algorithm: Multiple Linear Regression and XGBoost. [21]
- Output: Predicted crop yield (kg/acre).

Disease Detection.

- Input: Leaves images.
- Algorithm: Convolutional Neural Network (CNN).

• Output: Classification of plant disease (e.g. Early Blight, Leaf spot, Healthy).

3.4 Phase 4: Decision Support System

The visualizations of the processed data and model outputs were displayed on a dashboard (Grafana / Custom Web App) along with alerts and recommendations sent to farmers via a mobile app.

Some examples of recommendations included:

• "Soil moisture below threshold: Irrigate in next 6 hours."

• "Nitrogen deficiency detected: Apply 40kg/acre urea."

• "Leaf spot detected: Use fungicide spray."

IV.RESULTS

4.1 Experimental Design

The configuration and testing of the proposed system were carried out in a pilot farm setting (10 acres, wheat crop). An IoT device (ESP32 microcontrollers, DHT22 sensors, soil moisture sensors and soil pH sensors) were deployed across three plots in a field. To capture images of the leaves, a Raspberry Pi camera module was deployed. The data was subsequently sent to a cloud-based database (Firebase) via Wi-Fi at intervals of 10 minutes.[19][20]

• The programming languages used were Python and Arduino C.

• The following libraries were used: scikit-learn, TensorFlow, Pandas, NumPy.

• The visualization used in the system was a Grafana dashboard connected to the Firebase database.

• For machine learning models, Random Forest, Linear Regression and CNN were used.

4.2 Dataset

• Sensor Data: Collected over one cropping season (Kharif 2024).

• Plant Disease Images: PlantVillage dataset (54,000 leaf images, 14 crop species, 26 diseases).

• Yield Data: Historical wheat yield records (2015–2023) from government agriculture department.

4.3 Performance Evaluation

1. Irrigation Prediction

o Input features: soil moisture, air temperature, humidity, rainfall.

o Model: Random Forest Classifier.

o Result: Achieved 92% accuracy in predicting irrigation requirement.

o Example: System correctly identified 34 out of 37 irrigation events.

2. Crop Yield Estimation

o Model: Multiple Linear Regression + XGBoost.

o Metric: Mean Absolute Error (MAE).

o Result: MAE = 8% compared to actual yield.

o Example: Predicted yield of wheat = 2.3 tons/acre, actual = 2.5 tons/acre.

3. Disease Detection

o Model: Convolutional Neural Network (CNN).

o Training set: 80% of PlantVillage dataset.

o Testing set: 20% of dataset + field images.

o Result: **95% classification accuracy** on test images.

o Example: Correctly identified early blight and leaf spot in wheat leaves.

4.4 Visualization

The real-time monitoring dashboard (Grafana) effectively displayed the soil moisture, temperature, and humidity data to the farmer. Additionally, via mobile interface with access to crop data, growers viewed information on:

- Sensor readings updated every 10 minutes.
- Irrigation recommendations (e.g., "Irrigation needed within next 6 hours").
- Plant disease alerts (e.g., "Leaf spot detected, apply fungicide").

4.5 Discussion

This study presents the successful application of a data-centric agriculture system that integrates IoT-enabled sensing, cloud computing (cloud), and machine learning (ML) capabilities for smart farming. Our proposed architecture allowed for efficient real-time collection of soil, weather, and crop health data, and provided preprocessing and predictive models to produce actionable insight for farmers. Experimental validation demonstrated the proposed architecture was promising with irrigation prediction achieving 92% accuracy in reducing water waste, yield monitoring leading to only 8% prediction error improvements for better planning, and disease monitoring The result of the study reflects that the system has the capability to validate precision agriculture through accurate and on-time predictions and recommendations. The irrigation model reduced unnecessary water use (by ~18%), and the disease detections allowed for early intervention and minimize crop loss. The findings of this investigation demonstrate that the integration of IoT and machine learning can meaningfully improve grower decision making.

V.CONCLUSION

By using CNNs achieving 95% accuracy for intervention. The findings confirm that data-driven approaches enhance resource utilization, increase crop yields, and can improve sustainability through evidence-based recommendations related to the conditions in the field as opposed to just intuition-based recommendations. Future work can focus on introducing edge computing to provide farmers with a model that fast tracks the predictions right in the field, using blockchain for data sharing and transparency in the supply chain, scaling the system to multiple crops and larger farms, and using predictive weather models to support decision making. Overall, this study has demonstrated that data-driven agriculture can replace traditional agriculture with a precise, efficient, and sustainable practice; sustainable food security as a response to population growth and climate change are possible with data-driven practices.

REFERENCES

[1]. S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, “Big Data in Smart Farming – A review,” Agricultural Systems, vol. 153, pp. 69–80, 2017. doi: 10.1016/j.agsy.2017.01.023.

[2]. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” Sensors, vol. 18, no. 8, p. 2674, 2018. doi: 10.3390/s18082674.

[3]. A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018. doi: 10.1016/j.compag.2018.02.016.

[4]. T. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” Computers and Electronics in Agriculture, vol. 151, pp. 61–69, 2018. doi: 10.1016/j.compag.2018.05.012.

[5]. S. Patel and A. Patel, “Internet of Things-IoT: Definition, characteristics, architecture, enabling technologies, application & future challenges,” International Journal of Engineering Science and Computing, vol. 6, no. 5, pp. 6122–6131, 2016.

[6]. J. W. Jones, P. J. 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, “The DSSAT cropping system model,” European Journal of Agronomy, vol. 18, no. 3–4, pp. 235–265, 2003. doi: 10.1016/S1161-0301(02)00107-7.

[7]. A. Shinde and S. Kulkarni, “Crop disease detection using convolutional neural networks,” in Proc. IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018, pp. 1995–2000. doi: 10.1109/ICACCI.2018.8554932.

[8]. Mishra, A., Gupta, P., & Tewari, P. (2022). Global U-net with amalgamation of inception model and improved kernel variation for MRI brain image segmentation. Multimedia Tools and Applications, 81(16), 23339-23354.

[9]. Ramdoss, V. S., & Rajan, P. D. M. (2025). Evaluating the Effectiveness of APM Tools (Dynatrace, AppDynamics) in Real-Time Performance Monitoring. The Eastasouth Journal of Information System and Computer Science, 2(03), 399-402.

[10]. Chaturvedi, R. P., Mishra, A., Asthana, S., Parashar, M., & Nayyar, P. Embroilment of Deep Learning in Business Analytics for Sustainable Growth. Intelligent Business Analytics, 191-211.

[11]. Naphtali JH, Misra S, Wejin J, Agrawal A, Oluranti J. An intelligent hydroponic farm monitoring system using IoT. In Data, Engineering and Applications: Select Proceedings of IDEA 2021 2022 Oct 12 (pp. 409-420). Singapore: Springer Nature Singapore.

[12]. Joaquim MM, Kamble AW, Misra S, Badejo J, Agrawal A. IoT and machine learning based anomaly detection in WSN for a smart greenhouse. InData, Engineering and Applications: Select Proceedings of IDEA 2021 2022 Oct 12 (pp. 421-431). Singapore: Springer Nature Singapore.

[13]. Vijarania M, Udbhav M, Gupta S, Kumar R, Agarwal A. Global cost of living in different geographical areas using the concept of NLP. In Handbook of Research on Applications of AI, Digital Twin, and Internet of Things for Sustainable Development 2023 (pp. 419-436). IGI Global.

[14]. Gupta, S., Vijarania, M., Agarwal, A., Yadav, A., Mandadi, R. R., & Panday, S. (2024). Big Data Analytics in Healthcare Sector: Potential Strength and Challenges. In Advancement of Data Processing Methods for Artificial and Computing Intelligence (pp. 41-67). River Publishers.

[15]. Afah, D., Gautam, A., Misra, S., Agrawal, A., Damaševičius, R., & Maskeliūnas, R. (2021, February). Smartphones verification and identification by the use of fingerprint. In International Conference on Emerging Applications of Information Technology (pp. 365-373). Singapore: Springer Singapore.

[16]. Deore, H., Agrawal, A., Jaglan, V., Nagpal, P., & Sharma, M. M. (2020). A new approach for navigation and traffic signs indication using map integrated augmented reality for self-driving cars. Scalable Computing: Practice and Experience, 21(3), 441-450.

[17]. Vijarania M, Gupta S, Agrawal A, Misra S. Achieving sustainable development goals in cyber security using aiot for healthcare application. InArtificial Intelligence of Things for Achieving Sustainable Development Goals 2024 Mar 9 (pp. 207-231). Cham: Springer Nature Switzerland.

[18]. Singh A, Prakash N, Jain A. Chronic Diseases Prediction using two different pipelines TPOT and Genetic Algorithm based models: A

Comparative analysis. InProceedings of the 2024 9th International Conference on Machine Learning Technologies 2024 May 24 (pp. 175-180).

[19]. Singh A, Prakash N, Jain A. A comparative study of metaheuristic-based machine learning classifiers using non-parametric tests for the detection of COPD severity grade.

[20]. Singh A, Payal A. CAD diagnosis by predicting stenosis in arteries using data mining process. Intelligent Decision Technologies. 2021 Feb;15(1):59-68.

[21]. Singh A, Prakash N, Jain A. A review on prevalence of worldwide COPD situation. Proceedings of Data Analytics and Management: ICDAM 2022. 2023 Mar 25:391-405.

[22]. Dalal, S., Lilhore, U. K., Simaiya, S., Prakash, D., Yadav, S., Kumar, K., & Kaushik, A. (2026). GenAD-SM: optimized transformer-VAE model for precision anomaly detection for smart manufacturing in industry 5.0. Journal of Intelligent Manufacturing, 1-21. https://doi.org/10.1007/s10845-025-02764-5

[23]. Bhutani, M., Dalal, S., Alhussein, M., Lilhore, U. K., Aurangzeb, K., & Hussain, A. (2025). SAD-GAN: A Novel Secure Anomaly Detection Framework for Enhancing the Resilience of Cyber-Physical Systems. Cognitive Computation, 17.0(4), 127.

[24]. Dalal, S., Dahiya, N., Kundu, S., Verma, A., Devi, G., Ayadi, M., Dubale, M., & Hashmi, A. (2025). GAN-CSA: Enhanced Generative Adversarial Networks for Accurate Detection and Surgical Guidance in Skull Base Brain Metastases. International Journal of Computational Intelligence Systems, 18.0(1), 310.

[25]. Dalal, S., Lilhore, U. K., Faujdar, N., Simaiya, S., Agrawal, A., Rani, U., & Mohan, A. (2025). Enhancing thyroid disease prediction with improved XGBoost model and bias management techniques. Multimedia Tools and Applications, 84.0(16), 16757-16788.

[26]. Dalal, S., Lilhore, U. K., Seth, B., Radulescu, M., & Hamrioui, S. (2025). A Hybrid Model for Short-Term Energy Load Prediction Based on Transfer Learning with LightGBM for Smart Grids in Smart Energy Systems. Journal of Urban Technology, 32.0(1), 49-75.

[27]. Kaur, N., Mittal, A., Lilhore, U. K., Simaiya, S., Dalal, S., Saleem, K., & Ghith, E. S. (2025). Securing fog computing in healthcare with a zero-trust approach and blockchain. EURASIP Journal on Wireless Communications and Networking, 2025.0(1), 5.

[28]. Lilhore, U. K., Dalal, S., Radulescu, M., & Barbulescu, M. (2025). Smart grid stability prediction model using two-way attention based hybrid deep learning and MPSO. Energy Exploration & Exploitation, 43.0(1), 142-168.

[29]. Lilhore, U. K., Simaiya, S., & Dalal, S. (2025). 10 Hybrid Mathematical Optimization Techniques in AI. Math Optimization for Artificial Intelligence: Heuristic and Metaheuristic Methods for Robotics and Machine Learning, 2.0, 223.

[30]. Lilhore, U. K., Simaiya, S., Alhussein, M., Dalal, S., Aurangzeb, K., & Hussain, A. (2025). An Attention-Driven Hybrid Deep Neural Network for Enhanced Heart Disease Classification. Expert Systems, 42.0(2), e13791.

[31]. Lilhore, U. K., Simaiya, S., Dalal, S., & Faujdar, N. (2025). Revolutionizing air quality forecasting: Fusion of state-of-the-art deep learning models for precise classification. Urban Climate, 59.0, 102308.

[32]. Lilhore, U. K., Simaiya, S., Dalal, S., Alshuhail, A., & Almusharraf, A. (2025). A Post-Quantum Hybrid Encryption Framework for Securing Biometric Data in Consumer Electronics. IEEE Transactions on Consumer Electronics.

[33]. Lilhore, U. K., Simaiya, S., Dalal, S., Radulescu, M., & Balsalobre-Lorente, D. (2025). Intelligent waste sorting for sustainable environment: A hybrid deep learning and transfer learning model. Gondwana Research, 146.0, 252-266.

[34]. Malik, N., Kalonia, A., Dalal, S., & Le, D. N. (2025). Optimized XGBoost Hyper-Parameter Tuned Model with Krill Herd Algorithm (KHA) for Accurate Drinking Water Quality Prediction. SN Computer Science, 6.0(3), 263.

[35]. Ritika, R., Chhillar, R. S., Dalal, S., Moorthi, I., Dubale, M., & Hashmi, A. (2025). Enhanced heart disease diagnosis and management: A multi-phase framework leveraging deep learning and personalized nutrition. PLoS One, 20.0(10), e0334217.

[36]. Yadav, S., Sehrawat, H., Jaglan, V., Singh, S., Kantha, P., Goyal, P., & Dalal, S. (2025). A Novel Effective Forecasting Model Developed Using Ensemble Machine Learning For Early Prognosis of Asthma Attack and Risk Grade Analysis. Scalable Computing: Practice and Experience, 26.0(1), 398-414.

