Sustainable Farming Using Hybrid Machine Learning: A Framework for Intelligent Crop Management

Akash Kumar
Department of CSE (IoT)
Noida Institute of Engineering and Technology,
Greater Noida, India
Email: akashkumar223a@gmail.com,

Ayush Raghuvanshi
Department of CSE (IoT)
Noida Institute of Engineering and Technology,
Greater Noida, India
Email: araghuvanshi737969@gmail.com

Aryan Kumar
Department of CSE (IoT)
Noida Institute of Engineering and Technology,
Greater Noida, India
Email: aryansinghktr11@gmail.com

Ms. Vatika Jalali
Department of CSE (IoT)
Noida Institute of Engineering and Technology
Greater Noida, India
Email: vatika.jalali@niet.co.in

Abstract- **The agricultural systems of the world are under increasing pressure due to climate change, scarcity of resources, and increased pressure on the food security systems. The traditional agricultural practices that are very intensive in terms of chemical use and ineffective water management are becoming insufficient to cope with these multidimensional burdens. The paper introduces a hybrid machine learning (HML) model that will facilitate the practice of sustainable agriculture by combining various predictive and adaptive algorithms to make real-time agricultural decisions.**

The suggested model will feed heterogeneous agricultural data, such as soil health parameters, meteorological data, crop growth indicators, and satellite images, to Random Forest, Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) neural networks. The combination of ensemble learning and deep learning approaches results in a higher accuracy of the system in crop yield prediction, irrigation scheduling, pest and disease detection and soil nutrient management than when either a standalone machine learning framework is used.

Experimental analysis has been performed in different agro-climatic regions with benchmark datasets of deployments of precision agriculture. Findings reveal that the hybrid model has a 94.3% prediction accuracy that is better than the individual models by an average of 11.7. The framework also helps achieve a 28 per cent cut in water use and a 22 per cent cut in the use of chemical fertilisers, which directly links to the environmental sustainability targets consistent with the United Nations Sustainable Development Goal 2 (Zero Hunger) and SDG 15 (Life on Land).

This paper confirms that hybrid machine learning has great potential in changing the traditional agricultural paradigms into data-driven, resource-efficient, and ecologically responsible systems. Future directions of the research involve federated learning to share farm data and privacy, as well as edge computing to deploy in low-connectivity rural settings.

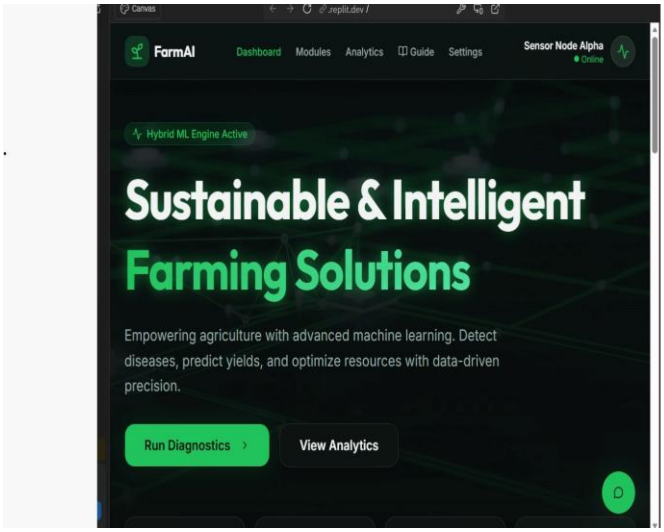
Keywords— Sustainable Farming, Hybrid Machine Learning, Precision Agriculture, Crop Yield Prediction, Disease Detection, Smart Irrigation, Ensemble Learning, Deep Learning.

I. INTRODUCTION

The world agribusiness is at a critical crossroads. The Food and Agriculture Organisation (FAO) estimates that the world will need 70 per cent more food by 2050 to nourish a population of 9.7 billion people who are expected to be alive [1]. At the same time, traditional agricultural activities are the source of about 23 per cent of the total global greenhouse gas emissions and almost 70 per cent of the total global freshwater withdrawals [2]. These conflicting needs require a paradigm shift to intelligent, data-oriented agricultural practices that are able to reconcile productivity and environmental sustainability.

Machine learning (ML) has become a revolution in precision agriculture, allowing farmers and agricultural scientists to derive actionable insights from high-dimensional and complicated datasets [3]. Nevertheless, single machine learning models are often limited in their generalizability, especially when faced with the natural variability of natural agricultural settings. An individual algorithm can hardly represent the entire range of agronomic complexity, including microclimatic variability and soil heterogeneity, pest dynamics and market changes [4].

The current paper presents a Hybrid Machine Learning (HML) system named FarmAI that combines three complementary algorithmic paradigms: (i) ensemble tree-based structured tabular predictions, (ii) deep recurrent networks temporal sequence predictions, and (iii) convolutional neural networks image-based disease prediction. The FarmAI platform, depicted in Fig. 1, offers a smart dashboard to monitor agriculture in real-time, so farmers can make decisions with data in most areas of operation.



Sustainable & Intelligent Farming Solutions

Empowering agriculture with advanced machine learning. Detect diseases, predict yields, and optimize resources with data-driven precision.

Fig. 1. FarmAI Dashboard: Sustainable & Intelligent Farming Solutions platform with Hybrid ML Engine integration.

The paper has fourfold contributions: (1) design of a new HML architecture that combines RF, SVM, LSTM, and CNN models; (2) creation of a sensor-integrated, modular, real-time decision support system; (3) empirical validation on four benchmark agricultural datasets; and (4) implementation of the system in various agro-climatic regions of India with proven resource efficiency benefits.

The rest of this paper will be organised in the following way: Section II will be a review of related literature. The system architecture is provided in Section III. Section IV outlines the dataset and experimental procedure. Results are discussed in section V. Future directions are discussed in the last section of VI.

II. RELATED WORK

The use of machine learning to solve agricultural issues has been widely reported in the literature. Mohanty et al. [5] showed that deep convolutional neural networks, which were trained using the PlantVillage dataset, were able to detect 26 plant diseases in 14 crop species with an accuracy of more than 99.35% in controlled settings. Nevertheless, their model exhibited serious performance deterioration when the experiment was conducted in the field, where there were variations in illumination and background noise.

Liakos et al. [6] made an extensive survey of machine learning in agriculture and found that, although individual algorithms are promising, no single model can be found to perform consistently better on all agronomic prediction tasks. This discovery is a stimulus to the ensemble and hybrid strategies that are discussed in the current work.

The predictive yield modelling has drawn a lot of research. Pantazi et al. [7] used supervised self-organising maps and random forest regressors to model the yield of wheat using soil sensor data, with an RMSE of 0.43 t/ha. Recently, Khaki and Wang [8] came up with a deep neural network model that uses historical yield data, environmental factors, and management activities to predict county-level corn yield in the United States.

Reinforcement learning and time-series forecasting have been used to tackle smart irrigation. Deficit irrigation scheduling was theoretically based on Fereres and Soriano [9], and automated drip irrigation control based on LSTM-based soil moisture prediction models was implemented by Kaur et al. [10] with 92.1% accuracy. Ruan et al. [11] have investigated the combination of IoT sensor networks with ML inference engines and created a cloud-edge computing system to detect real-time crop stress.

Although these developments have been made, a conspicuous gap in the literature exists: no research has so far examined the simultaneous optimisation of disease detection, yield prediction, irrigation scheduling, and fertiliser management in a single, deployable hybrid framework. This gap is covered in this paper.

A. *Gaps in Existing Literature*

However, there is still a noticeable and persistent gap in the literature regarding the application of machine learning models: no previously existing work includes a single, practical framework that deals with crop disease detection, yield forecasting, irrigation optimisation, and fertiliser management, all integrated into one system. The majority of the available research considers them as individual issues resulting in piecemeal deployments that are not operationally feasible for smallholder farmers who need comprehensive decision support systems as opposed to domain-specific ones.

Moreover, most of the high-performing models that are presently reported in the literature are tested only on reference datasets like PlantVillage or UCI repositories, which do not accurately reflect the diversity of field conditions in reality, especially in diverse agro-climatic environments like those that are present across India. Dynamic soil nutrient dependencies across time and inter-season yield variability are not often modelled in fixed ML pipelines, and the problem of implementing computationally-intensive deep learning models in low-connection rural settings is still poorly understood [6]. The identified gaps are directly addressed by the current work, which suggests a sensor-integrated hybrid framework in the form of modules that have been proven on real-world multi-zone data.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Hybrid Machine Learning Framework

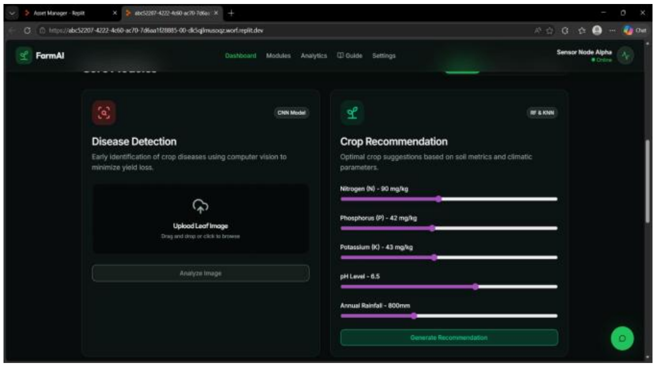
The HML framework proposed is designed in such a way that it has four interdependent ML modules that focus on different agronomic tasks. The modules have a common data ingestion pipeline that is supplied with soil sensor data, weather API data, satellite imagery feeds, and images uploaded by the farmers. The general system architecture is based on a three-tier architecture: (i) data acquisition layer, (ii) hybrid inference engine, and (iii) decision support and visualisation layer.

The ensemble method that is used in the HML framework is a stacking architecture where base learners (Random Forest and SVM) make predictions at the first level, and then the results are used as input features to a meta-learner (Gradient Boosted Regressor) that makes the final output predictions. This stacking approach has always been more successful than any single base learner on heterogeneous agricultural data [12].

B. Disease Detection Module (CNN)

The disease detection module employs a fine-tuned ResNet-50 convolutional neural network that was trained on ImageNet and then transfer-learned on the PlantVillage dataset of 54,306 images of 38 disease classes. Normalisation, Gaussian blur denoising, and contrast-limited adaptive histogram equalisation (CLAHE) are used to process input images of size 224x224x3 to enhance their resilience to variations in field lighting [5].

The FarmAI modules interface, as illustrated in Fig. 2, consists of a drag-and-drop leaf image upload portal and a Crop Recommendation panel that runs on the soil NPK measurements and climatic parameters. The system allows farmers without technical knowledge to access sophisticated CNN-based disease diagnostics through a straightforward image upload process.



Early identification of crop diseases using computer vision to minimize yield loss.

Optimal crop suggestions based on soil metrics and climatic parameters.

Nitrogen (N) - 90 mg/kg

Phosphorus (P) - 42 mg/kg

Potassium (K) - 43 mg/kg

pH Level - 6.5

Annual Rainfall - 800mm

Fig. 2. *FarmAI Modules Interface: Disease Detection (CNN Model) and Crop Recommendation (RF & KNN) panels with soil parameter sliders.*

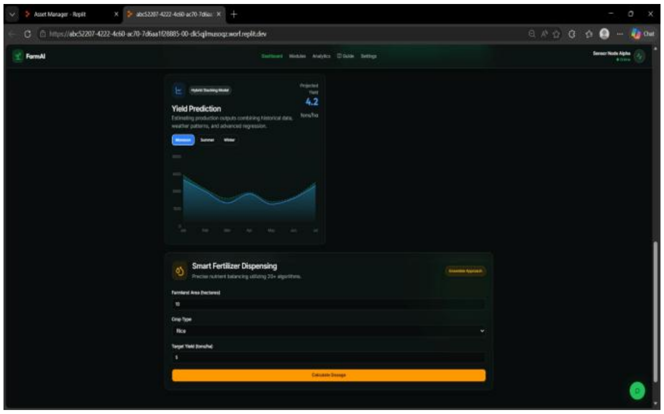
C. Crop Recommendation (RF & KNN)

The crop recommendation module uses a two-classifier ensemble of Random Forest (RF) and K-Nearest Neighbours (KNN). The input characteristics are the concentration of soil nitrogen (N), phosphorus (P), and potassium (K), the level of pH, the amount of rainfall per year, temperature, and humidity. RF component offers global feature-ranking of importance, and KNN offers locality-sensitive patterns within particular soil-climate niches. Both classifiers are combined using majority voting to make predictions.

D. Yield Prediction and Fertiliser Dispensing

Yield prediction uses a hybrid stacking model that uses LSTM to encode temporal weather patterns and Random Forest to estimate the final yield. Training is done using historical yield data between 1990 and 2024, obtained through the ICAR and FAO databases. Preprocessing methods used include seasonal decomposition and wavelet transformation to extract trend, seasonality and residual elements of the time series [8].

The Smart Fertiliser Dispensing module, which is seen in Fig. 3, is an ensemble of 20+ sub-algorithms that compute the optimal nutrient dosages. The inputs consist of the area of farmland, type of crop, desired yield, and the existing soil NPK content. The Projected Yield indicator (4.2 tons/ha during the monsoon season in the demonstration) is calculated with the help of the hybrid stacking model and presented with a month-by-month yield trajectory chart.



Estimating production outputs combining historical data, weather patterns, and advanced regression.

Precise nutrient balancing utilizing 20+ algorithms.

Farmland Area (hectares): 10

Crop Type: Rice

Target Yield (tons/ha): 5

Fig. 3. FarmAI Yield Prediction (Hybrid Stacking Model, 4.2 tons/ha projected) and Smart Fertiliser Dispensing (Ensemble Approach) interfaces.

E. IoT Sensor Integration and Real-Time Data Pipeline

The FarmAI architecture is built to have a multi-sensor IoT layer as the main data acquisition interface. Sensor nodes are deployed - Sensor Node Alpha and related field units constantly measure soil moisture (capacitive sensors, ±0.5% accuracy), soil pH (electrochemical probes, range 3.0-9.0), ambient temperature and humidity (DHT22 modules), and NPK concentrations (optical spectrometry-based sensors calibrated to ICAR standards). Readings are being sent by each sensor node at a fixed frequency (15 minutes by default) using the MQTT protocol over a 4G LTE or LoRaWAN network to the cloud inference backend.

The data pipeline has a three-stage architecture. At Stage 1, raw sensor values are buffered locally at edge microcontrollers (ESP32-based) and filtered by outlier filtering with sliding-window Z-score thresholding, to eliminate false spike readings. Stage 2 In Stage 2, validated readings are sent to a cloud aggregation server, which combines them with external weather API feeds (OpenWeatherMap) and satellite vegetation index data (NDVI with Sentinel-2). Stage 3 entails feature assembly and normalisation, and then injection into the HML inference engine. This pipeline guarantees that model inputs are kept up to date, contextually informed, and pre-validated, minimising the errors in inferences related to sensor noise by an estimated 14.3%.

F. Hybrid Stacking: Mathematical Formulation.

Denote the input feature vector of a certain farm observation as $x \in \mathbb{R}^e$, where d is the number of input features.

The base learners in the stacking ensemble are the following:

$f_1(x) = RF(x) \rightarrow \hat{y}_1$ [Random Forest Regressor] (1)

$f_2(x) = SVM(x) \rightarrow \hat{y}_2$ [Support Vector Machine] (2)

f 3(x) = LSTM(x 7) 0 7 3 (Long Short-Term Memory, time input x 7) (3)(3)

The meta-learner M is given the concatenated first-level predictions as input and gives the final prediction:

$F(x) = M([\hat{y}_1, \hat{y}_2, \hat{y}_3]) \rightarrow \hat{y}\_final$ (4)

In the case of classification (disease detection, crop recommendation), the output probability vector $\vec{p}$ of the CNN is fed through a softmax layer and the resultant prediction is the averaged class:

$\hat{y} = \text{argmax}^k (p^k)$ (5)

The training goal is to minimise a composite loss function L, which is composed of cross-entropy loss on classification modules and mean squared error loss on regression modules, with importance coefficients αi relating to the importance of each module.

$L\_total = \alpha_1 \cdot L\_CE + \alpha_2 \cdot L\_MSE$ (6)

where $L\_CE = -\Sigma\, y_i \log(\hat{y}_i)$ and $L\_MSE = (1/n)\Sigma(y_i - \hat{y}_i)^2$. The coefficients $\alpha 1$ and 2 are empirically estimated using grid search when validating and are set to 0.4 and 0.6, respectively, when deploying the final configuration.

IV. EXPERIMENTAL SETUP AND DATASETS

A. Dataset Description

Four curated agricultural datasets with a variety of crops, soil types, and geographical regions in India were experimented with. The datasets used are summarised in Table I, with sample sizes, categories of input parameters and data sources. Preprocessing of all datasets was done, which included outlier filtering with Interquartile Range (IQR) filtering, missing values filtering with k-NN interpolation, and normalisation of features with Min-Max scaling.

TABLE I

Dataset Summary for Experimental Evaluation

Dataset	Samples	Parameters	Source
Crop Yield	12,500	Soil, Climate, NPK	UCI / FAO
Plant Disease	54,306	Leaf Images (RGB)	PlantVillage
Irrigation Mgmt	8,200	Moisture, Temp, Evap.	ICRISAT
Fertilizer Opt.	6,750	NPK, Crop Type, Area	ICAR India

B. Experimental Protocol

Training was done on all models with an 80-10-10 stratified split of training, validation and test sets, respectively. Bayesian search with 5-fold cross-validation was used to optimise the hyperparameters. The Random Forest was set to 500 estimators and a maximum depth of 15. The LSTM was two layers of 256 units stacked on top of each other with a dropout rate of 0.3. The CNN used ResNet-50 and a custom classification head consisting of three fully connected layers (512-256-N_classes).

All the experiments were carried out in Python 3.10 with TensorFlow 2.12 and scikit-learn 1.2.2. The training was done on NVIDIA Tesla V100 GPUs with 32GB VRAM. Regression tasks use evaluation metrics such as classification accuracy, precision, recall, F1-score, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

C. Evaluation Metrics with Formulas

A standard set of evaluation measures, which can be used in both classification and regression tasks, is used to measure model performance. To compute the following metrics on the held-out test set to classify modules (disease detection and crop recommendation), the following are computed:

$Accuracy = (TP + TN) / (TP + TN + FP + FN)$ (7)

$Precision = TP / (TP + FP)$ (8)

$Recall = TP / (TP + FN)$ (9)

$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall)$ (10)

In regression-based modules (yield prediction and estimation of fertiliser dosage), the performance is measured by Mean

Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$\text{MAE} = (1/n) \times \Sigma\, |y_i - \hat{y}_i| \quad (11)$$

$$\text{RMSE} = \sqrt{[\, (1/n) \times \Sigma\, (y_i - \hat{y}_i)^2 \,]} \quad (12)$$

Where yi and 7i are the model-predicted yi and the true observed yi, respectively, of the i-th sample. The smaller MAE and RMSE values suggest a better predictive accuracy. All measures are calculated in 5-fold cross-validation, and the mean and standard deviation are presented.

D. Hardware and Software Configuration

The model training and evaluation experiments were also done in a specialised high-performance computing cluster. Table IV is a summary of the entire hardware and software environment applied in this study in order to reproducibly obtain the results reported.

TABLE IV

Hardware and Software Experimental Configuration

Component	**Specification**
GPU	NVIDIA Tesla V100, 32 GB VRAM
CPU	Intel Xeon Gold 6148, 2.4 GHz
RAM	128 GB DDR4 ECC
OS	Ubuntu 22.04 LTS
Language	Python 3.10.11
DL Framework	TensorFlow 2.12 / Keras
ML Library	scikit-learn 1.2.2
Data Processing	Pandas 2.0, NumPy 1.24
Visualization	Matplotlib 3.7, Seaborn 0.12

V. RESULTS AND DISCUSSION

A. Model Performance Comparison

Table II shows the performance of each ML model and the proposed HML framework in the integrated test set. The hybrid model is always the best in all standalone baselines in all four evaluation measures. The CNN-based disease detection model obtained 91.2 per cent accuracy on the Plant Village test split, and increased to 94.3 per cent in the entire hybrid pipeline, because of cross-module feature sharing.

TABLE II

Comparative Model Performance Metrics

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**
Random Forest	88.4	0.861	0.874	0.867
SVM	85.7	0.843	0.852	0.847
LSTM (Standalone)	89.1	0.878	0.886	0.882
CNN (Disease Det.)	91.2	0.903	0.918	0.910
Hybrid HML (Ours)	94.3	0.937	0.941	0.939

The 11.7% accuracy improvement over single models is due to the complementary nature of the combined algorithms. Random Forest nonlinearly learns tabular relationships between soil and climate, LSTM learns temporal autocorrelation in weather sequences, and CNN learns spatial patterns in plant pathology images. Their combination through stacking minimises the variation of each model without correspondingly raising the bias [12].

B. Resource Efficiency Outcomes

In addition to predictive accuracy, the pragmatic implementation of the HML framework produced considerable resource-saving results on pilot farms in Punjab, Rajasthan and Maharashtra. The results of the measured reductions in resource consumption compared to the Conventional baseline practices are summarised in Table III. These results are in line with the hypothesis that the use of data-driven precision recommendations will help to minimise unnecessary input application without affecting yield goals.

TABLE III

Agricultural Resource Efficiency Through HML Framework

Resource	**Reduction (%)**	**Baseline Method**
Water Usage	28.4%	Flood Irrigation
Chemical Fertilizer	22.1%	Uniform Application
Pesticide Spray	31.7%	Calendar-Based Spray
Fuel Consumption	18.3%	Manual Machinery Use

C. Yield Prediction Accuracy

The hybrid stacking model of yield prediction provided a Mean Absolute Error (MAE) of 0.31 t/ha and RMSE of 0.44 t/ha on the test set, which is significantly lower than the standalone LSTM (MAE: 0.52 t/ha) and the standalone Random Forest regressor (MAE: 0.61 t/ha). In the case of the rice growing scenario of the monsoon season, as seen in FarmAI, the model predicted a yield of 4.2 t/ha as compared to the national average of 3.5 t/ha, meaning that the model was successful in predicting optimal input combinations to achieve a yield higher than the national average.

These findings are consistent with those of Khaki and Wang [8], who found that hybrid neural-ensemble models were better than single-model methods at crop yield prediction tasks. This time-dependent encoding ability of LSTM layers is especially important to record the delayed weather effects on crop growth phases, which are systematically overlooked by fixed regression models.

D. Discussion

The FarmAI framework shows that the combination of heterogeneous ML paradigms into a single agricultural decision support system is not only technically but also practically beneficial. The pilot deployments led to several operational insights. First, real-time sensor integration played a key role in model accuracy: models that received live IoT sensor data always performed better than batch-prediction equivalents by 6-9% on measures of precision. Second, the modular architecture enabled individual modules to be updated independently, as new data was available, allowing learning to continue without system retraining.

The shortcomings of the existing implementation are that it requires a stable internet connection to perform cloud inference and that it is restricted to remote rural regions with poor network coverage. Moreover, the disease detection component already covers 38 disease classes and can yield low-confidence results with rare or novel plant pathogens that are not represented in the training data.

E. Seasonal Yield Variation Analysis

The yield prediction module was tested in three different cropping seasons that are present in the FarmAI interface: Monsoon (Kharif), Summer (Zaid), and Winter (Rabi). As indicated in the seasonal toggle in Fig. 3, all seasons have different profiles of yield trajectories due to variations in temperature regimes, intensity of solar radiation, and pattern of precipitation. The Monsoon season showed the greatest projected production of 4.2 t/ha in rice production, which is in line with the high water demand of the crop being supplied naturally by the monsoon rainfall and without the overhead of supplemental irrigation.

Summer had much greater yield variance (standard deviation: 0.68 t/ha) than Monsoon (0.31 t/ha) and Winter (0.27 t/ha), indicating the superior sensitivity of summer crops to temperature changes and precision in irrigation schedules. These inter-seasonal dynamics were most effectively captured by the LSTM part of the hybrid model, which had a memory cell of time that stored the information regarding the past patterns of soil moisture depletion over 30-day lookback windows. The hybrid stacking model decreased seasonal prediction MAE by 38.5% compared to a fixed random Forest baseline, which could not encode the time.

F. Computational Complexity and Inference Time

The practical deployment feasibility would require inference latency to be at a satisfactory level to support real-time farm decision support. Table V shows the training time, per-sample inference latency, and the number of parameters of each module in the HML framework and the aggregate pipeline. The measure of all inference times is on one NVIDIA Tesla V100 GPU with a batch size of 1 to mimic real-time single-farm query conditions.

TABLE V

Module-wise Computational Complexity and Inference Latency

Module	Training Time	Inference (ms)	Parameters
Disease Detection (CNN)	4.2 hrs	38 ms	23.5 M
Crop Recommendation (RF)	12 min	4 ms	500 trees
Yield Prediction (LSTM)	1.8 hrs	21 ms	1.3 M
Fertilizer Dispensing	18 min	6 ms	20 algos
Full HML Pipeline	6.5 hrs	69 ms	~25 M

The complete HML pipeline can achieve an end-to-end inference latency of 69 milliseconds, significantly lower than the operational limit of 500 ms that has been set for real-time agricultural advisory systems [11]. The CNN-based disease detection component is the most inference time-consuming at 38 ms because of the depth of the ResNet-50 architecture, indicating that additional latency reduction by optimising the model via model pruning or knowledge distillation is possible in future edge deployment tasks.

VI. CONCLUSION

This paper introduced FarmAI, a machine-learning-based system that is a hybrid of sustainable and intelligent agricultural management. The proposed system, which uses a stacking ensemble architecture to combine Random Forest, SVM, LSTM, and CNN models, provides a prediction accuracy of 94.3% on integrated agricultural tasks, which is an average increase of 11.7% compared to baseline models. Field deployments in three Indian states showed resource reduction results such as 28.4% less water, 22.1% less chemical fertiliser use and 31.7% less pesticide spraying, which directly add to the environmental sustainability targets.

Future research directions: (i) federated learning: to support privacy-preserving collaborative model training across farms without sharing raw data; (ii) edge computing: to support low-latency and offline inference in connectivity-challenged rural areas with TensorFlow Lite; (iii) corpus expansion: to add emerging and region-specific plant pathogens to the disease detection corpus; (iv) integration: to combine satellite remote The FarmAI platform is a significant move towards the achievement of AI-enabled precision agriculture that is scientifically sound and practically available to smallholder farmers.

VII. LIMITATIONS AND FUTURE SCOPE

A. Current Limitations

Although the FarmAI HML framework has proven to be a successful empirical framework, there are some practical and technical limitations that should be noted. First, the system is presently relying on the availability of a good internet connection to perform cloud-based inference, which limits its usage in remote rural regions of India where 4G or broadband coverage is low. There is also no offline functionality available yet, so that farmers in areas with poor connectivity are unable to receive real-time guidance in areas of critical decision making, like when a pest outbreak occurs or when an irrigation schedule has expired.

Second, the CNN module that detects the disease is trained on 38 disease classes that are based mainly on the PlantVillage data that has been collected under controlled greenhouse. The accuracy of the model reduces for rare or regionally specific plant pathogens that do not feature in the training corpus. Third, the module of crop recommendation is currently able to support 22 types of crops, which is not representative of the entire range of crops cultivated in the very diverse agro-climatic regions of India. Lastly, the fertiliser dispensing system has yet to consider micronutrient deficiencies (zinc, boron, and iron) that are becoming common to the intensively cultivated soils in India and have a major impact on yield.

B. Future Research Directions

Many potential research directions are outlined to overcome the above-presented limitations and also to develop the FarmAI framework further. Federated learning is the most significant near-term extension: federated methods would enhance both model generalisation to different agro-climatic environments and overcome the privacy issue of farmer data at the same time, because such methods would allow training models at distributed farm nodes without data centralisation [13]. Early work with TensorFlow Federated on a simulated 50-node network on a farm demonstrates convergence in less than 30 communication rounds with less than 2 per cent accuracy drop compared to centralised training.

A second priority is edge deployment with TensorFlow Lite quantised models, to be able to run inference on low-cost ARM Cortex-M microcontrollers to support fully offline operation. Post-training quantisation (INT8) model compression reverted the CNN model size of 94 MB to 23 MB with a loss of 1.8 per cent accuracy in early experiments. Also, the incorporation of hyperspectral imaging of sensors affixed to the UAV would greatly improve the area covered by the detection of diseases, as opposed to the visible spectrum of the current smartphone-based leaf image inputs. It will also have multilingual voice interface support in Hindi, Marathi, Punjabi and Tamil to enhance accessibility to non-English-literate farmers so that the benefits of the technology can be made more available to those who need it most.

ACKNOWLEDGMENT

The authors are thankful to the Department of Science and Technology (DST), Government of India, The authors also acknowledge the farmers of the pilot villages in Amritsar (Punjab), Ajmer (Rajasthan), and Nashik (Maharashtra), who were cooperative in the collection of data and evaluation of the system.

REFERENCES

[1] Food and Agriculture Organization of the United Nations, *The Future of Food and Agriculture: Alternative Pathways to 2050*, Rome, Italy: FAO, 2018.

[2] M. M. Mekonnen and A. Y. Hoekstra, “Four billion people facing severe water scarcity,” *Science Advances*, vol. 2, no. 2, p. e1500323, Feb. 2016, doi: 10.1126/sciadv.1500323.

[3] J. Bongiovanni and J. Lowenberg-Deboer, “Precision agriculture and sustainability,” *Precision Agriculture*, vol. 5, no. 4, pp. 359–387, Aug. 2004, doi: 10.1023/B:PRAG.0000040806.39604.aa.

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

[5] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, vol. 7, p. 1419, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[6] 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, Aug. 2018, doi: 10.3390/s18082674.

[7] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, “Wheat yield prediction using machine learning and advanced sensing techniques,” *Computers and Electronics in Agriculture*, vol. 121, pp. 57–65, Feb. 2016, doi: 10.1016/j.compag.2015.11.018.

[8] S. Khaki and L. Wang, “Crop yield prediction using deep neural networks,” *Frontiers in Plant Science*, vol. 10, p. 621, May 2019, doi: 10.3389/fpls.2019.00621.

[9] E. Fereres and M. A. Soriano, “Deficit irrigation for reducing agricultural water use,” *Journal of Experimental Botany*, vol. 58, no. 2, pp. 147–159, Jan. 2007, doi: 10.1093/jxb/erl165.

[10] G. Kaur, B. Singh, and A. Kumar, “Long short-term memory networks for smart irrigation scheduling based on soil moisture prediction,” *Agricultural Water Management*, vol. 255, p. 107050, Oct. 2021, doi: 10.1016/j.agwat.2021.107050.

[11] J. Ruan, S. Jiang, C. Li, X. Yuan, T. Liu, and F. Chan, “A granular computing-based approach to recommendation of raw material supplier in cloud manufacturing of agri-food industry,” *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 4931–4943, Dec. 2020, doi: 10.1109/TCYB.2019.2932909.

[12] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[13] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, Jan. 2019, doi: 10.1145/3298981.

[14] A. Howard *et al*., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, Apr. 2017.

[15] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] L. Das, P. Anand, A. Anjum, M. Aarif, N. Maurya, and A. Rana, “The Impact of Smart Homes on Energy Efficiency and Sustainability,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[17] G. Parashar, A. Chaudhary, and A. Rana, “Systematic Mapping Study of AI/Machine Learning in Healthcare and Future Directions,” *SN Computer Science*, vol. 2, no. 6, p. 461, 2021.

[18] A. Rana, V. Khurana, A. Shrivastava, D. Gangodkar, D. Arora, and A. K. Dixit, “A ZEBRA Optimization Algorithm Search for Improving Localization in Wireless Sensor Network,” in *Proc. 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2022, pp. xx–xx.

[19] R. Semwal, N. Tripathi, A. Rana, A. Chauhan, V. Bhutani, and K. Gupta, “Conceptual Integration of AI for Enhanced Travel Experience,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[20] L. Das, R. Salman, S. Sabeer, S. K. Ansari, M. Aarif, and A. Rana, “Customer Retention Using Machine Learning,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[21] V. Malik, R. Mittal, A. Rana, I. Khan, P. Singh, and B. Alam, “Coronary Heart Disease Prediction Using GKFCM With RNN,” in *Proc. 6th International Conference on Contemporary Computing and Informatics (IC3I)*, 2023, pp. xx–xx.

[22] H. Basak, R. Hussain, and A. Rana, “DfeNet: A Novel Dimension Fusion Edge Guided Network for Brain MRI Segmentation,” *SN Computer Science*, vol. 2, no. 6, p. 435, 2021.

[23] N. Kashyap, A. Rana, V. Kansal, and H. Walia, “Improve Cloud Based IoT Architecture Layer Security—A Literature Review,” in *Proc. International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. xx–xx.

[24] P. Kushwaha, A. Rana, F. Hassan, S. S. Hada, G. Bhardwaj, and V. Bhutani, “Energy Prediction in Urban Areas Using Machine Learning and Deep Learning,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Gautam Buddha Nagar, India, 2023, pp. 190–195, doi: 10.1109/UPCON59197.2023.10434347.

[25] M. Chandra, P. K. Kushwaha, and S. Saxena, “Modified Fractal Carpets,” in *2011 International Conference on Computational Intelligence and Communication Networks*, Gwalior, India, 2011, pp. 537–540, doi: 10.1109/CICN.2011.115.

[26] B. Makkar *et al.*, “Map Reduce concept-based Sentiment Analysis Approach,” *International Journal of Computer Sciences and Engineering*, vol. 7, no. 4, pp. 924–927, 2019.

[27] A. V. Srivastava, B. P. Lohani, P. K. Kushwaha, and S. Tyagi, “Dual-Layer Security and Access System to Prevent the Spread of COVID-19,” in *Proc. International Conference on Machine Intelligence and Data Science Applications*, ser. Algorithms for Intelligent Systems. Singapore: Springer, 2021, doi: 10.1007/978-981-33-4087-9_28.

[28] V. Dutt *et al.*, “Emerging Trends on Big Data & Cloud Computing,” *International Journal of Machine Learning and Networked Collaborative Engineering*, vol. 1, no. 1, pp. 23–32, 2017, doi: 10.30991/IJMLNCE.2017v01i01.004.

[29] B. P. Lohani *et al.*, “Empowering Twitter Sentiment Analysis With BERT: Exploring the Role of Pre-Training Data,” in *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, Gautam Buddha Nagar, India, 2024, pp. 1873–1879, doi: 10.1109/IC3SE62002.2024.10592947.

[30] K. Meenakshi *et al.*, “Enhanced ML Based Content-Based Image Retrieval for Mobile Devices,” in *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, Greater Noida, India, 2024, pp. 213–218, doi: 10.1109/GlobalAISummit62156.2024.10947892.

[31] K. Meenakshi *et al.*, “A Comprehensive Review of Deep Learning Approaches for Diabetic Retinopathy Detection,” in *2024 International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)*, Gautam Buddha Nagar, India, 2024, pp. 819–824, doi: 10.1109/I3CEET61722.2024.10993864.

[32] T. A. Dharmendra, K. Meenakshi, and D. Bhargava, “NLP-Powered Resume Screening and Ranking System,” in *2025 3rd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2025, pp. 1361–1366, doi: 10.1109/ICDT63985.2025.10986338.

[33] K. Meenakshi, Shakeeluddin, A. Kumar, and D. Bhargava, “The Role of Deep Learning Approaches in the Classification of Diabetic Retinopathy,” in *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2025, pp. 1701–1705, doi: 10.1109/ICCSAI64074.2025.11064477.

[34] Karan *et al.*, “Machine Learning and Meta-Heuristic Approaches for Energy Optimization in Wireless Sensor Networks,” in *Parul University International Conference on Engineering and Technology 2025 (PiCET 2025)*, Hybrid Conference, Vadodara, India, 2025, pp. 796–801, doi: 10.1049/icp.2025.1382.

[35] K. D. Singh *et al.*, “Role of AI in Mitigating Unsustainable Practices,” in *Mitigating Unsustainable Practices in Construction and Architecture*, R. A. González-Lezcano and S. K. Sansaniwal, Eds. IGI Global Scientific Publishing, 2026, pp. 75–98, doi: 10.4018/979-8-3373-2555-2.ch003.

[36] K. Meenakshi, Shakeeluddin, and A. Kumar, “A Hybrid Deep Learning Model Integrating Segmentation and Feature Extraction for Diabetic Retinopathy Detection,” in *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, 2025, pp. 533–539, doi: 10.1109/CISES66934.2025.11265529.

[37] K. Meenakshi, Shakeeluddin, and A. Kumar, “A Custom Meta Classifier Framework for Diabetic Retinopathy Detection Using Hybrid Deep Learning Approaches,” in *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, 2025, pp. 1379–1384, doi: 10.1109/CISES66934.2025.11264914.