Optimizing Vegetable Fertigation with IoT and ML in Controlled Ecosystems

Geetha Radhakrishnan[1]*, Roy Stephen[1] and V. S. Santhosh Mithra[2]

[1]*Kerala Agricultural University, Thrissur 680 656, Kerala, India*
[2]*ICAR-CTCRI, Thiruvananthapuram- 695 017. Kerala, India*

Received on12 March 2024; received in revised form 13 July 2024, accepted 02 September 2024.

Abstract
This study presents an approach for optimizing fertigation in a controlled setting, utilizing Internet of Things (IoT) and Machine Learning (ML) to enhance water and fertilizer management for vegetable crops, under rain shelter conditions. Using IoT sensors, data are relayed to a cloud system, and analyzed by RF, CNN, RNN and LSTM models. These models predict essential NPK for crops, enabling automated fertigation to maintain optimal growth conditions. This also promotes resource efficiency and environmental conservation. This approach ensures that crops receive adequate water and nutrients, safeguarding yield quality, conserving water, and preserving soil health, thereby marking a pivotal advancement in sustainable farming.

Keywords: Artificial Intelligence (AI), Hi-Tech agriculture, Internet of Things (IoT), Machine Learning (ML), Precision Farming, Sensors, Sustainable agriculture

Introduction

According to the Ministry of Finance (2023), the expanding Indian economy has gained global prominence over the past decade. The Ministry of Agriculture & Farmers Welfare (2021) and the Ministry of Statistics and Programme Implementation, report that agriculture contributes 17% to the national GDP, playing a vital role in the economy. Pattanayak and Mallick (2017) indicate that while approximately 70% of Indians are farmers, the agricultural sector faces challenges from population growth, urbanization, climate change, and resource depletion. Shakoor and Ullah (2024) observe that many farmers still rely on traditional methods, outdated irrigation, excessive fertilizer use, and struggle with unpredictable weather, limiting productivity. Key factors like soil quality, moisture, weather, and disease management are essential for a successful harvest (Magdoff, 2001). Neglecting these aspects can lead to inefficient resource use and unhealthy crops, with excessive fertilizers and chemicals harming soil composition (Kurnjawati et al., 2023).

To address these issues, introducing smart farming technologies is critical for informed decision-making. Soil temperature monitoring is crucial for agriculture, impacting plant growth, seed germination, nitrification, soil moisture, and nutrient availability (Galezewski et al., 2021). Astija (2020) shows that soil pH, ranging from 0 (acidic) to 14 (alkaline), is measured with sensors to assess nutrient levels and detect unwanted chemicals, noting that tomato plants develop best when pH is near 7. Neina (2019) affirms that pH monitoring helps maintain this level through fertilizer application. Navyashree (2023) emphasizes the role of water in nutrient transport within plants; adequate water circulation supports plant rigidity and health,

* Author for Correspondences: Phone:9745425056; Email:geetha.ramesh@kau.in

even with evapotranspiration. Soil moisture properties, such as field capacity, Permanent Wilting Point (PWP), and available water, vary by soil type (Liu and Ma, 2024). Electrical conductivity, a measure of soil salt content, serves as an indicator of nutrient and water availability, influencing crop suitability (Blaschek et al., 2019). High soil salinity can hinder plant growth (Gondek et al., 2020), making salt level monitoring essential for crop selection, irrigation, and soil management. Monitoring NPK levels is also vital as these macronutrients are required in varying amounts across crops (Isreal and Yonas, 2021). Micronutrients are needed in smaller quantities, usually grams per hectare, and regular NPK checks ensure proper nutrient supply for optimal growth and yield (Feyissa et al., 2022).

This paper introduces smart fertigation system for control through a User Interface (UI), enabling data aggregation, visualization, and analysis in the cloud. Data analytics employ Machine Learning (ML) and Deep Learning (DL) techniques, focusing on models such as Random Forest (RF), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). The digital twin integrates automated irrigation and fertigation, thereby minimising manual labour, promoting optimal fertiliser and water use.

Materials and Methods

Identifying the Water Requirements of Crop

Plants of the tomato variety "Vellayani Vijay" were raised in grow trays in a naturally ventilated green house for 30 days. During this period, the relative humidity spanned between 71% and 99% and the mean temperature ranged from 24°C to 30°C. Uniformly sized seedlings were then transplanted to separate earthen pots having a top diameter of 35 cm and then filled with soil, well decomposed farmyard manure, and sand in the proportion 2:1:1. Vermicompost was also added to enhance the soil fertility and water retention capabilities, and kept in a rain shelter of 250 square meter area. Four pots were maintained equidistantly with a gap of 10 centimeters from the edge of one pot to the edge of the next one. The plants were irrigated and the soil moisture in each pot was maintained at 70-80%, 60-70%, 50-60%, and 40-50% respectively. This was achieved using the gravimetric method, where the soil moisture in the pots was measured periodically and the water was added to reach and maintain the desired percentage. In the gravimetric method, the weight of the pot, dry soil, and wet soil are measured and the moisture per cent of soil in each pot is computed using the formula:

$$soil\ moisture\ content\ (\%) = \frac{wgt\ wet\ soil - wgt\ dry\ soil}{wgt\ wet\ soil\ in\ pot - wgt\ dry\ soil\ in\ pot} \times 100 \quad \text{— (1)}$$

Eighty per cent (80%) of soil moisture, seemed to be the maximum water-holding capacity of the soil in each pot. After this percentage, the water seeped out of the pot. Sensors were also used to measure the soil moisture. The resulting values were compared to the computed soil moisture values to confirm accuracy. This approach ensured precise control over the soil moisture levels, facilitating a comprehensive study of the effects of varied water availability on plant growth. Moisture levels were computed at the edge server, and the controller within the rain shelter activated the actuators accordingly. Plant growth parameters pertaining to plant height and the number of leaves from each pot were taken manually at fifteen-day intervals until the 70th day.The total weight of fruits from each plant was also computed.

Assessment of pH and EC in the soil

The determination of nitrogen (N), phosphorus (P), and potassium (K) values in soil samples involves a series of meticulous steps, ensuring accuracy and precision in the results. Five zones within the rain shelter are chosen to ensure a representative sampling. Based on the soil sampling methods explained by Dawson and Knowles, (2018), soil samples were collected from varying depths, since nutrient concentration can change with depth. The samples are air-dried to eliminate moisture content, and then sieved to remove coarse materials like stones, roots, and leaves, ensuring uniformity for

accurate measurements. Ten grams of finely sieved soil are weighed, and 25 milliliters of distilled water is added. This soil-to-water ratio facilitates nutrient and parameter extraction from the soil. The soil-water mixture is placed in a shaker to create a homogeneous solution and enhance the extraction of soluble nutrients.

The suspension is stirred at regular 30-minute intervals. This regular agitation ensures a consistent extraction of nutrients from the soil particles into the water. Once the soil particles have adequately settled after the final stirring, the electrical conductivity of the soil is measured using an EC meter. This measurement displaysthe salinity or total dissolved salts in the soil.A calibrated pH meter is dipped into the soilsuspension to measure the pH value. Soil pH is an indicator of soil acidity or alkalinity and can influence the availability of nutrients to plants.

Computation of Soil N,P,K values

Colorimetric Analysis was performed to ascertain the availability of phosphorous (P). Phosphorus was extracted from the soil using an appropriate extraction solution. The extracted P reacts with a reagent to produce a colored compound. The intensity of this color is directly proportional to the concentration of phosphorus in the sample. A colorimeter or spectrophotometer is used to measure the intensity of the developed color. The phosphorus concentration in the sample is determined by comparing the reading with a calibration curve created from known phosphorus standards (Ekukinam et al., 2014).

Potassium was extracted from the soil using the Flame Photometry method with a neutral 1N ammonium acetate solution. This solution effectively extracts exchangeable potassium ions from the soil. The extracted solution is introduced into the flame of a photometer, where potassium vaporizes and emits light at a characteristic wavelength. The photometer measures this light intensity, and the potassium concentration is determined by comparing it to a standard calibration curve.

Walkley and Black Method was used to assess the availability of Organic Carbon in the soil. This method oxidizes, soil organic carbon using a mixture of potassium dichromate ($K_2 Cr_2 O_7$) and sulfuric acid, releasing carbon dioxide (CO_2) proportional to the organic carbon content. After the oxidation process, the excess chromic acid is titrated using ferrous ammonium sulphate to determine the amount of chromic acid reduced. This gives the amount of carbon present in the soil. Organic carbon percentage is then calculated based on the titration values.

The Kjeldahl Method is used to compute nitrogen (N) content in the soil. Soil is digested with concentrated sulphuric acid, which converts all nitrogen forms to ammonium (NH_4^+). The digested solution is then distilled with sodium hydroxide, releasing ammonia. The ammonia is trapped in a boric acid solution and titrated with a standard acid solution. Nitrogen content is calculated based on the titration values (Jackson, 1973).

Setting the field for automated irrigation and fertigation

'Vellayani Vijay' tomato plants were grown and transplanted into earthen pots with a top circumference of 35 cm. The pots were placed 10 cm apart in an equidistant arrangement. Each plant received 1.2 liters of water on alternate days through drip irrigation to sustain soil moisture within the 50–60% range. To maintain humidity between 71% and 99% within the shelter, mist irrigation was applied.

Installation of sensors for data collection

Pamungkas et al. (2023) has explained the importance and effectivity of using soil sensors to measure soil nutrients and moisture. Through their study, Le et al. (2023) examines FDR-based sensors to monitor soil moisture, nutrient levels, and their influence on crop health and find it a better option.

Prabu et al., 2023 discusses the integration of corrosion-resistant sensors for soil nutrient monitoring in varying soil conditions. Keeping these factors in mind, for a more nuanced understanding of soil nitrogen, phosphorous, and potassium, RS485 wireless sensors which adopt the Frequency Domain Reflectometry (FDR) measurement method were inserted 4 inches deep into the soil. This sensor has high-density epoxy resin filled between the probe and body to ensure high waterproofing. The probes are made of austenitic 316 stainless steel to make it salt-alkali corrosion resistant, rust-proof, and electrolytic resistant. The sensor has a highly sensitive chip that consumes low power, to provide stable signals.

Figure 1 depicts multiple sensors with independent identification numbers attached in a bus network over a long distance, and these multiple sensors use a data logger and a single port (RS232/serial). The main sensors used in this device are the *7 in 1 Soil Integrated Sensor* (1) for monitoring nitrogen, phosphorous, and potassium (NPK), electrical conductivity, pH, and moisture; the *Water flow sensor* (2) that gives the volume of water being used for irrigation at any given instant, and *atmospheric parameter sensors* (3) to measure temperature and humidity with RS485 module. The real-time data collected from the point of transplanting the seedlings are uploaded to a cloud-based data base at 5-minute intervals 24/7, for a period of ninety days. The sensors are connected to the controller, an edge device.

Workflow in the automation process

The processes involved in this automation are depicted in Figure 2. The sensors collect data and store them in their own memory space until requested by the controller. The data collected by one of the sensors is in the form of electrical conductivity and then gets converted to numerical values. Another sensor senses the passage of water through the pipes in the form of frequency pulses.This is then used by the controller to compute the volume of water delivered. Temperature and humidity sensors collect the related data and push it to the controller through the I^2T (IsquareT) communication protocol. The controller makes data requests to the sensors at regular intervals and collects only valid data. Basic levels of

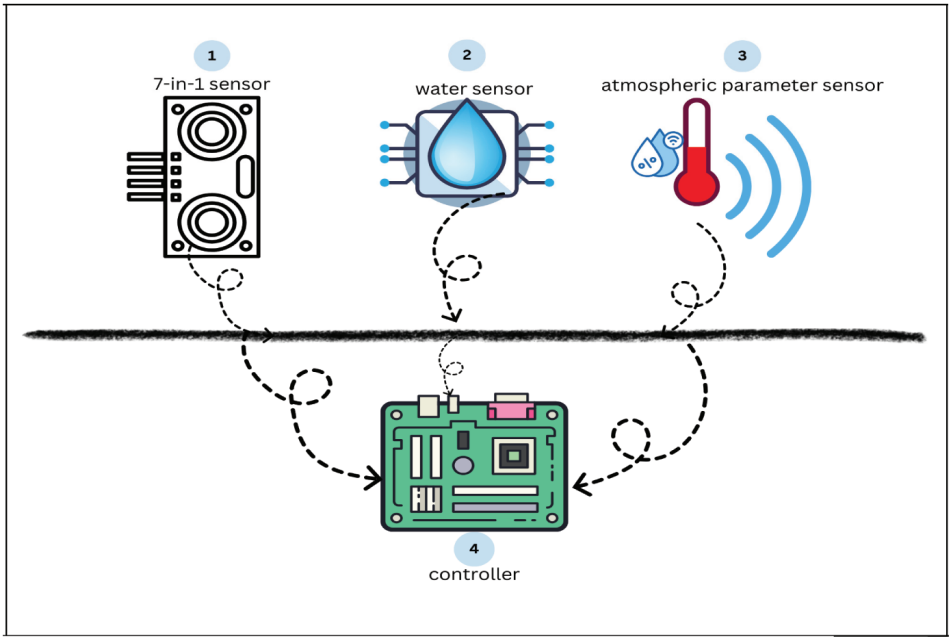

Figure 1: Sensor-Microcontroller connectivity (Edge device)

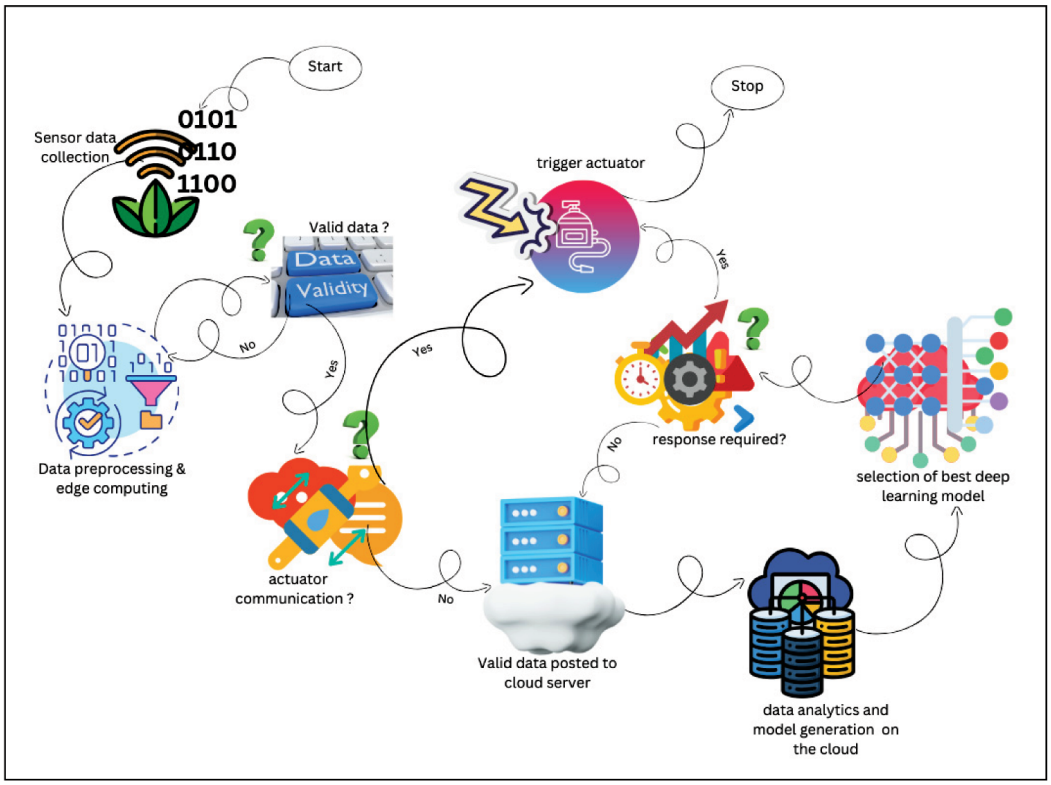

Figure 2. Process Flow Diagram of IoT Fertigation System

computations, data validation, and pre-processing are done within the controller. Most of the actuator activations are done based on decisions at the edge level. Only specific valid data are pushed to the database server in the cloud. A combination of deep learning techniques is applied to the data to obtain model stacks and recommendations. Frequent analytics at regular intervals and dynamic corrections are made to the models until the best values for optimum irrigation and fertigation recommendation are obtained for crops in a particular soil and field type. Based on the requests made by the users, the results from the models are implemented through actuators that are under the supervision of the controllers.

Implementation of IoT and sensors in the field

As depicted in Fig. 3 multiple sensors with independent identification numbers are attached in a bus network over along distance, and these multiple sensors use a single port (RS232/serial), and data logger. The RS485 sensors transmit signals over twisted cable/dual lines and follow the MODBUS master/slave protocol of communication, wherein the master node initiates the communication. All nodes connected will receive the call but only the slave node that is meant to will respond. The sensor layer in the Edge device communicates locally with the Modbus server in Gateway in the networking layer via Modbus protocol. It then pushes the data to the cloud server and communicates through the MQTT pubsub-OASIS protocol. The data server in the processing layer collects and stores the data being transmitted. Rule engines, controls, decision-making, and visualization details are stored in the application server.The database server and application server continuously communicate with each other. The user publishes a notification related to the information requested. The MQTT server puts out this request in the gateway through the MQTT client. When this information is available with the MQTT server in

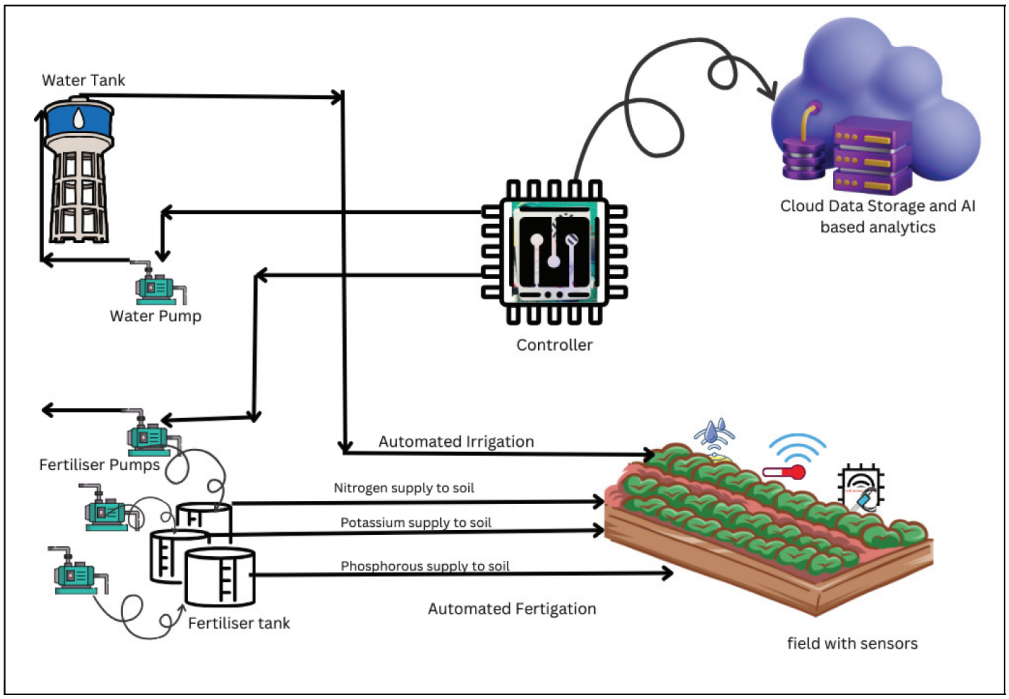

Figure 3. Process Flow Diagram of IoT Fertigation System

the cloud, the user is notified and the actuators are activated as per the rules set in the rule engine. Codes have been written using the Python programming language to perform data analytics on the real-time data that is pushed into the cloud server. AI-based analyses are performed on the real-time data and a model isgenerated, based on which the quantity of water and nutrients is determined for crops in a field at any given location. Data is transmitted from the edge server to the cloud through wireless network systems. The power supply required for the continuous monitoring of the soil parameters by the sensors is provided by 1KVA sonic power solar PCU with MPPT, 260 W solar panel and 200AH tubular batteries.

Data Processing and Analytics

Random Forest (RF): A Random Forest model is an ensemble machine-learning technique that constructs multiple decision trees during training and outputs an average prediction of individual trees for regression tasks or a mode (majority vote) for classification tasks. By Building several trees and averaging results, the model reduces the risk of over fitting and improves generalization. They are especially useful for small databases because they can capture complex relationships without requiring a massive amount of data, and they help prevent over fitting which often occurs with limited data sets (Simon et al., 2023).

Convoluted Neural Networks (CNN): CNNs can be adapted for time series data analysis by treating sequences as one-dimensional "images", capturing local temporal dependencies through convolution operations. This approach benefits from the ability of CNN to automatically and adaptively learn spatial hierarchies of features from the data. However, for very short sequences in small databases, the model may overfit, and CNN might require more data preprocessing compared to traditional time series methods (Subeesh and Mehta, 2021)

Recurrent Neural Network (RNN): These are inherently designed to handle sequential data, capturing long-term dependencies and temporal

dynamics, thereby making them more suited for time series than CNN. They are optimal for tasks where past information significantly influences future predictions, such as stock price prediction. However, RNNs can be computationally intensive and might suffer from long training times with large databases, and are prone to issues like vanishing gradients, which can make the training of deep RNNs challenging (Sampath and Sumithira, 2022; Hewamalage et al., 2021).

Long Short-Term Memory (LSTM): This is a type of RNN designed to capture long-term dependencies and avoid the vanishing gradient problem which is typical of the RNN. While CNNs specialize in spatial hierarchies and are ideal image data, LSTMs excel in understanding sequences and temporal hierarchies, making them more suited for time series and natural language processing tasks. Compared to basic RNNs, LSTMs can remember patterns over longer sequences, offering better performance for tasks with extended temporal patterns (van Houdt et al., 2020). However, LSTMs can be more computationally intensive than RNNs and CNNs.

The evaluation metrics are taken as Mean Square Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), and MSE for Mean Squared Error. This evaluates the accuracy of a regression model by computing the average of the squared between actual and predicted values. In formula terms, for data points, if *yi* denotes the real value and is its corresponding prediction, the MSE is given by:

$$MSE = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n} \quad \text{----------(2)}$$

Essentially, it quantifies how close a model's predictions are to the actual outcomes, with lower values signifying a better fit.

RMSE:RMSE, which stands for Root Mean Squared Error, is a widely used metricin regression analysis that measures the model's prediction accuracy. Essentially, it represents the standard deviation of the residuals (or prediction errors). By calculating the square root of the Mean Squared Error (MSE), it quantifies how much, on average, the model predictions deviate from the actual values. Mathematically, for n data points, where *yi* is the actual value and $\hat{y}$ i, is the predicted value for the i[th] observation, RMSE is computed as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad \text{----------(3)}$$

A lower RMSE value indicates a better fit of the model to the data, as it means the errors between the predicted and actual values are, on average, smaller. Conversely, a higher RMSE suggests larger errors and potentially a less accurate model.

Mean Absolute Error (MAE) is a metric used in regression analysis to evaluate a model's accuracy by computing the average of the absolute differences between the actual and predicted values. Its primary purpose is to provide a straight forward and easily interpretable measure of prediction error without squaring, emphasizing all errors uniformly. However, a disadvantage of MAE is that it does not highlight large errors, as it treats all errors equally, potentially missing significant outliers or deviations. The formula for Mean Absolute Error is given by:

$$MAE = \sum_{i=}^{n} \frac{|y_i - \hat{y}|}{n} \quad \text{----------(4)}$$

where:
n is the total number of observations
yi is the actual value for the *i*[th] observation
$\hat{y}$ is the predicted value for the *i*[th] observation
$|y_i - \hat{y}|$represents the absolute difference between the actual and predicted values for each observation

R-Squared: R^2 is a statistical measure that represents the proportion of the variance for the dependent variable that is explained by independent variables in a regression model. It provide same as ure of how well the observed outcomes are replicated by the model R^2 provides a measure of the strength of the relationship between the model and the response variable (Chicco et al., 2021). It offers an indication of the "goodness of fit" of the model. An R^2 value

of 1 indicates that the regression predictions perfectly fit the data, while a value of 0 indicates no fit.The formula for R^2 is as given below:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad \text{--------} \quad (5)$$

Where:

SS_{res} is the sum of squares of residuals (the differences between observed and predicted values SS_{tot} is the total sum of squares (the variance of the dependent variable) Since accuracy increases when the R^2 value gets closer to1, the results obtained from each of the models were compared and the best one was selected, based on the R^2 values

Results and Discussion

In Figure 4, the x-axis represents soil moisture levels in pots, measured as a percentage (%), while the y-axis indicates plant height in inches and the number of leaves. The graph compares soil moisture ranges of 40-50%, 50-60%, 60-70%, and 70-80%, with data collected from four pots (Pot 1, Pot 2, Pot 3, Pot 4). Each moisture range displays corresponding plant height and leaf count.

Observations reveal that maintaining soil moisture at 70-80% causes water seepage from the pots, requiring irrigation only once every two days. At 60-70% soil moisture, alternate-day irrigation suffices without seepage. Similarly, 50-60% soil moisture also supports alternate-day irrigation, with a volume of 1.2 litres per pot. When the pH value is maintained between 6.3 to 7.0, in moisture ranges of 70-80% and 60-70%, plants achieved similar heights, whereas, at 50-60% and 40-50%, the height of the plant decreases. The number of leaves in the plants were inversely proportional to the decrease in soil moisture. Although leaf numbers were maximum at 40-50% soil moisture, it exhibited unhealthy characteristics, bore fruits of the lowest weight, and raised soil pH to 8.0.

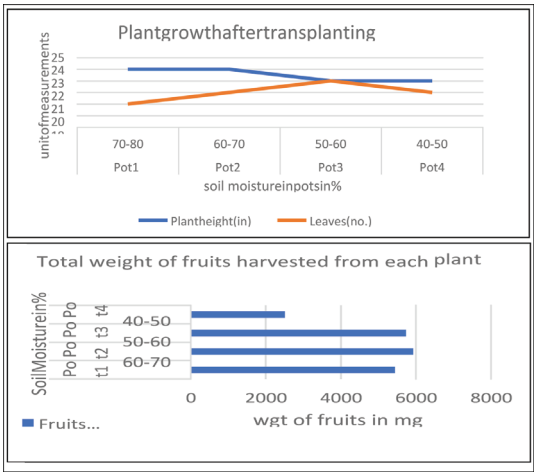

Figure 4. Computing soil moisture requirement

Tomato plants yielded the heaviest fruits at 60-70% soil moisture, followed by plants at 50-60%. A table accompanying the graph details the total fruit weight harvested from each plant across different moisture levels, with measurements provided in milligrams (mg).

This study aims to identify the optimal conditions for plant growth and yield through controlled irrigation. Soil pH is maintained between 6.3 and 7.0 to protect the root system, ensuring effective uptake of essential nutrients. In acidic soils, nutrient absorption slows, impeding tomato plant growth (Wang et al., 2000). Conversely, when pH shifts to the alkaline range, the soil becomes more susceptible to disease (Ibrahim et al., 2020). Therefore, maintaining a near-neutral pH is essential. The subsequent experiments focus on maintaining 50-60% soil moisture, which promotes efficient water use, keeps pH within the ideal range, and supports a healthy harvest.

Somasundaram (2018), discusses the role of CSV format for structured data ingestion on cloud analytics platforms. Hence, in the subsequent experiments, data is stored as a comma-separated values (CSV) file on cloud platforms. Collected at 30-minute intervals, the data is concatenated and integrated at the end of each day to ensure a continuous dataset (Orter et al., 2019) and to preserve the temporal context (Koh et al., 2021). Preprocessing is applied to this data to enhance result accuracy (Maharana et al., 2022) before

analytical procedures are conducted.

Outliers are managed using interquartile range analysis, which captures the middle 50% of the dataset, making it resistant to extreme values (Vinutha et al., 2018). Missing values, often due to machine failure, maintenance, or human error, are filled using interpolation to prevent biased interpretations (Noor et al., 2013). The data is transformed through normalization and standardization. The final dataset, containing N, P, and K values, can exhibit linear or non-linear patterns (Whetton et al., 2021). Soil N, P, and K levels are influenced by various factors with non-linear interdependencies, resulting in a complex time-series dataset (Qi et al., 2017). This refined dataset is subsequently used for testing and training with models such as Random Forest (RF), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM).

After training and testing, the evaluation of nitrogen predictions in Table 1 shows that the Random Forest (RF) model achieves the highest R^2 value (0.99) and the lowest Mean Absolute Error (MAE, 0.16) and Root Mean Square Error (RMSE, 0.29), indicating superior performance in terms of error metrics. In contrast, both CNN and RNN models display higher error rates, each with R^2 values of 0.96. The LSTM model shows the highest Mean Squared Error (MSE, 0.97) and RMSE (0.99), suggesting it may be less suitable for nitrogen prediction in this context.

Table 2 presents potassium evaluation results, where RF again demonstrates the lowest MAE (1.09) and RMSE (2.05) with a high R^2 of 0.97, making it the most accurate model for potassium prediction. The CNN model yields significantly higher errors across all metrics, with the lowest R^2 (0.88), indicating suboptimal performance. RNN and LSTM models, while showing elevated error values, maintain higher R^2 values of 0.94 and 0.95, respectively.

In Table 3, RF outperforms other models for Phosphorus prediction, with the lowest MSE (0.13), MAE (0.26), and RMSE (0.36), along with the highest R^2 (0.99). While CNN and RNN models show moderate performance with higher errors and R^2 values of 0.99 and 0.97, respectively, the LSTM model again exhibits the highest errors among all models.

Although the RF model is generally preferred for

Table1. Evaluation metrics for nitrogen

Model	MSE	MAE	RMSE	R^2
RF	0.71	0.16	0.29	0.99
CNN	0.69	0.70	0.83	0.96
RNN	0.66	0.62	0.82	0.95
LSTM	0.97	0.64	0.99	0.95

Table 2. Evaluation metrics for potassium

Model	MSE	MAE	RMSE	R^2
RF	4.21	1.09	2.05	0.97
CNN	18.71	3.04	4.32	0.88
RNN	7.27	1.88	2.70	0.94
LSTM	7.84	2.32	2.80	0.95

Table 3. Evaluation metrics for phosphorous

Model	MSE	MAE	RMSE	R^2
RF	0.13	0.26	0.36	0.99
CNN	1.34	0.92	0.96	0.99
RNN	1.03	0.77	1.01	0.97
LSTM	1.44	0.94	1.2	0.97

Table 4. Evaluation metrics for nitrogen

Model	MSE	MAE	RMSE	R^2
LSTM	0.20	0.20	0.44	0.99

Table 5. Evaluation metrics for potassium

Model	MSE	MAE	RMSE	R^2
LSTM	4.87	1.22	2.20	0.96

Table 6. Evaluation metrics for phosphorous

Model	MSE	MAE	RMSE	R^2
LSTM	0.41	0.38	0.64	0.99

AH : Ampere Hour; AI: Artificial Intelligence; CSV:Comma Separated Values; CNN: Convoluted Neural Network; DL: Deep Learning; EC: Electrical Conductivity; FDR: Frequency Domain Reflectometry; GDP: Gross Domestic Product IoT: Internet of Things; KVA : Kilo Volt Ampere; LSTM: Long Short-Term Memory; MAE: Mean Absolute Error; ML: Machine Language; MODBUS: MODular data BUS; MPPT: Maximum Power Point Tracking; MQTT: Message Queuing Telemetry Transport; MSE: Mean Square Error, NPK : Nitrogen, Phosphorous, Potassium; PCUpH: Power Conditioning Unit potential of Hydrogen ; PWP: Permanent Wilting Point; RF: Random Forest; RMSE: Root Mean Square Error ; RNN: Recurrent Neural Network; R^2 : R Squared; UI: User Interface; W: Watt

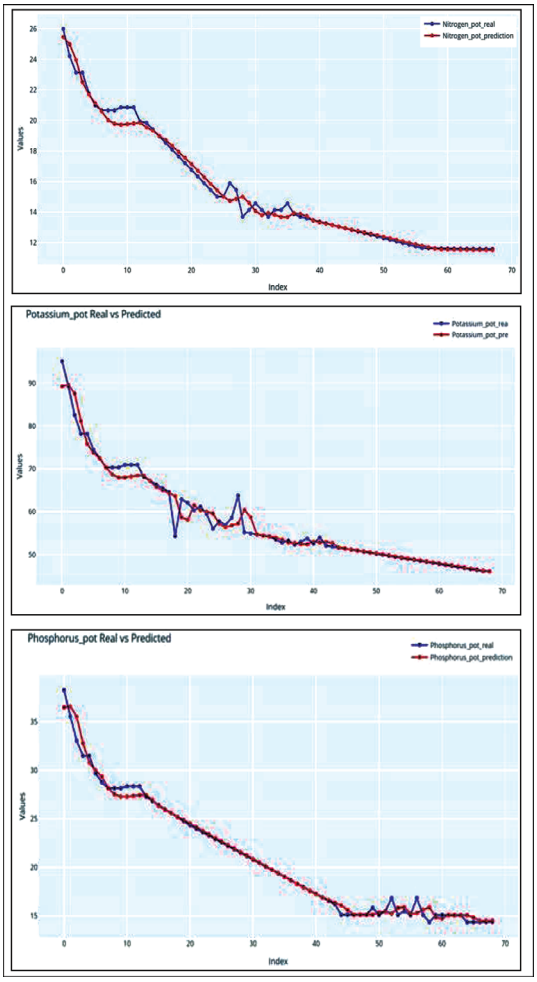

Figure 5. Graph plotting the actual and predicted values of NPK to be used for the crop.

a. Actual value and predicted value of nitrogen using the LSTM model.
b. Actual value and predicted value of phosphorus using the LSTM model.
c. Actual value and predicted value of potassium using the LSTM 0model.

non-time-series data, it can be adapted for one-step forecasting in time-series contexts when datasets are small. RF provides R^2 values of 0.99 for nitrogen, 0.97 for potassium, and 0.99 for phosphorus, indicating robust predictions for these nutrients. However, as database size grows, RF may risk model instability (Tyralis et al., 2017).

Numerous studies have demonstrated CNN efficacy with spatial data, while RNN is commonly applied to time-series data, though it suffers from vanishing gradients Sevin et al, 2017). Research by Liu et al. (2021) and Chandra et al. (2021) underscores LSTM effectiveness in time-series forecasting. LSTM accuracy often improves with adjusted epoch sizes; however, overly large epochs can lead to overfitting. Per Tensor Flow guidelines, epochs should be adjusted until accuracy stabilizes. The previous LSTM model employed 70 epochs with 32 batches. Fine-tuning this to 90 epochs with the same batch size yields improved metrics across Tables 4, 5, and 6, achieving R^2 values of 0.99 for nitrogen, 0.96 for potassium, and 0.99 for Phosphorus.

This optimized LSTM model with 90 epochs and 32 batches shows markedly better performance for nitrogen, potassium, and phosphorus predictions compared to the earlier LSTM model results in Tables 1, 2, and 3. Consequently, the fine-tuned LSTM model (illustrated in Figure 4) may be the preferred choice for forecasting fertilizer needs in the next cultivation cycle of tomato crops.

Conclusion

The study aimed to determine the optimal soil moisture level for the best plant growth and fruit yield. The data presented in the graph and tables offer insights into how the different soil moisture levels affect the various growth parameters such as plant height leaf number and fruit weight. The graph and table collectively indicate that a soil moisture range of 60-70% is optimal for plant growth. This range consistently shows better results in terms of plant height and leaf number.

Plants in pots with soil moisture levels of 60-70% exhibit the greatest height, suggesting that this moisture range is ideal for vertical growth. Similarly, the number of leaves is high in the same 60-70% soil moisture range, indicating robust vegetative growth under these conditions. The total weight of fruits harvested is highest in pots with 70-80% soil moisture. Although slightly higher moisture levels might decrease vegetative growth,

they seem to favor fruit production. Lower soil moisture levels of 40-50% tend to result in shorter plants with lower fruit yield. Conversely, overly high moisture levels, above 80% are not covered in the data but might potentially lead to issues such as root rot. The results suggest that maintaining soil moisture within the 60-70% range is generally beneficial for both vegetative growth and fruit production. For maximizing fruit yield specifically, slightly higher moisture levels of 70-80% may be more advantageous. These findings can guide agricultural practices, ensuring that crops receive optimal amounts of water for healthy growth and maximum yield.

The evaluation of data using various machine learning models to predict nitrogen, potassium, and phosphorus levels reveals that RF consistently outperforms other models across all metrics, with the lowest error rates and highest R^2 values, indicating high prediction accuracy. Conversely, CNN demonstrates higher error metrics, rendering it less effective for these predictions. The performances of RNN and LSTM models vary, with certain LSTM models showing metric-specific improvements. Previous studies have noted inaccuracies in RF predictions as the database scales up. RNN models, while capable of time-series analysis, encounter the vanishing gradient problem, limiting their ability to learn patterns over extended data sequences. In contrast, adjustments to LSTM models have produced competitive results in subsequent evaluations, suggesting potential for refinement. Consequently, the LSTM model emerges as the most suitable choice for nutrient level predictions in this study.

The fertilizer schedule predicted here should be implemented in the next crop cycle. This automation process for irrigation and fertigation has been conducted in red soil specifically for tomato crops. Additional experiments are required to apply this approach to the same crop in different soil types and also to other crops across various soils. These experiments should continue until a substantial database is built to refine and enhance prediction accuracy.

Acknowledgements

This manuscript is composed based on the preliminary findings from the ongoing project titled "Development of Smart Nursery Management System Using IoT and ML Techniques." This project is currently being executed with the support of the Kerala State Planning Board.

References

Astija A. (2020). Soil pH influences the development of tomato root organ (*Solanum lycopersicum* L.). Eurasian Journal of BioSciences, 14:6903-6908

Blaschek, M., Roudier, P., Poggio, M., Hedley C.B. (2019). Prediction of soil available water-holding capacity from visible near-infra red reflectance spectra. Scientific Reports, 9(1):12833. doi: 10.1038/s41598-019-49226-6.

Chandra R., Jain A., Singh Chauhan D. (2022). Deep learning via LSTM models for COVID-19 infection forecasting in India. PLoS ONE, 17(1): e0262708. https://doi.org/10.1371/journal.pone.0262708

Chicco D, Warrens MJ, Jurman G. (2021). The coefficient of determination R squared is “more informative than SMAPE, MAE, MAPE, MSE and RMSE inregression an alysis evaluation. PeerJComputSci., 7(1):e623.doi:10.7717/peerj-cs.623.

Dawson A. and Knowles O. (2018). Current Soil Sampling Methods- A Review, Proceedings Fertiliser and Lime Research Conference: Farm Environmental Planning - Science, Policy and Practice At: Palmerston North, New Zealand, pp. 1-11.

Ekukinam,E.U., Iwara A.I., and Gani B.S. (2014). Evaluation of phosphorous and exchangeable base status of soil under rubber plantation of different ages in south easternNigeria.Open Sci. J.Bio.Sci, Bio.Eng. 1(2):19-22.

Feyissa, T., Zhao, S., Ma, H., Duan Z., Zhang W. (2022). Optimizing plant density and balancing NPK inputs in combination with innovative fertilizer product for sustainable maize production in North China Plain. Sci Rep, 12:10279. https://doi.org/10.1038/s41598-022-13736-7

Galezewski L., Jaskulska I., Jaskulski, D. Arkadiusz

Lewandowski A., Szypolowska A., Wilczek A. and Szczepanczyk M. (2021). Analysis of the need for soil moisture, salinity and temperature sensing in agriculture: a case study in Poland. Scientific Reports, 11:16660. https://doi.org/10.1038/s41598-021- 96182-1

Gondek M., Weindorf D. C., Thiel C., Kleinheinz G. (2020). Soluble Salts in Compost and Their Effects on Soil and Plants: A Review, Compost Science & Utilization, 28:2, 59-75. doi:10.1080/1065657X. 2020.1772906

Hewamalage H, Bergmeir C., Bandara K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1)388- 427. https://doi.org/10.1016/j.ijforecast.2020.06.008,

Ibrahim Y.E., ElKomy, M.H., Balabel,N.M. et al. (2020). Saline and alkalines oil stress results in enhanced susceptibility to and severity in tomato pithnecrosis when inoculated with either *Pseudomonas corrugata* and/or *P. fluorescens*. J Plant Pathol102:849â•“856. https://doi.org/10.1007/s42161-020-00544-z

Isreal Z. and Yonas R. (2021). Review on the role of soil macronutrient (NPK) on the improvement, yield and quality of agronomic crops. Journal of Agriculture and Food Research, 9(1) :7-11. https://doi.org/ 10.26765/DRJAFS 23284767.

Jackson, M.L. (1973). Soil Chemical Analysis. Prentice Hall of India Pvt. Ltd., NewDelhi, 498.

Kurnjawati A., Toth G., Ylivainio K., Toth Z. (2023). Opportunities and challenges of bio-based fertilizers utilization for improvings oil health. Organic Agriculture, 13:335-350. https://doi.org/10.1007/ s13165-023-00432-7

Koh B.H.D., Lim C.L.P., Rahimi H., Woo W.L., Gao B. (2021). Deep Temporal Convolution Network for Time Series Classification. Sensors (Basel), 21(2):603. doi: 10.3390/s21020603.

Le M.T., Pham C.D., Nguyen T.P.T., Nyugen T.L., Nyugen Q.C., Hoang N. B., Nghiem L.D. (2023). Wireless Powered Moisture Sensors for Smart Agriculture and Pollution Prevention: Opportunities, Challenges, and Future Outlook.Curr Pollution Rep, 9:646-659. https://doi.org/10.1007/s40726-023-00286-3

Liu L. and Ma X. (2024). Prediction of Soil Field Capacity and Permanent Wilting Point Using Accessible Parameters by Machine Learning. AgriEngineering, 6(3):2592-2611. https://doi.org/ 10.3390/agriengineering6030151

LiuZ., Zhu Z., Gao J. and Xu C. (2021). Forecast Methods for Time Series Data: A Survey. IEEE Access, 9:91896-91912. doi: 10.1109/ACCESS. 2021. 3091162.

Maharana K., Mondal S., Nemade B. (2022). A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 3(1):91-99. https://doi.org/10.1016/j.gltp.2022.04.020.

Magdoff F. (2001). Concept, components, and strategies of soil health in agroecosystems. Journal of Nematology, 33(4):169-72.

Ministry of Statistics and Programme Implementation (2021). Sector- wise GDP of India https://www. statisticstimes.com/economy/country/ india-gdp-sectorwise. php

Ministry of Agriculture & Farmers Welfare (2021). Contribution of Agriculture Sector towards GDP, https://www.pib.gov.in/Press Release Page.aspx? PRID=1741942 Ministry of Finance (2023). Economic Survey 2022-23: Highlights https:// www.pib.gov.in/Press Release I frame Page.aspx? PRID=1894929

Navyashree, R. (2023). Water relations and uptake in crops. Crop Physiology a Collaborative Insights: Kumari A and Gali S., 1:190648

Neina, D. (2019). The role of soil pH in plant nutrition and soil remediation. Applied and Environmental Soil Science, 5794869. https://doi.org/10.1155/2019/ 5794869

Noor N.M., Ramli N.A., Abdullah M.M.A.B. (2013). Filling Missing Data Using Interpolation Methods: Study on the Effect of Fitting Distribution. Key Engineering Materials, 594-595, 889-895. https:// doi.org/10.4028/www.scientific.net/KEM.594-595.889.

Orter S., Ravi D.K., Singh NB, Vogl F., Taylor W.R., Ignasiak K. N. (2019). A method to concatenate multiple short time series for evaluating dynamic behaviour during walking. *PLoSONE*, 14(6): e0218594. https://doi.org/10.1371/journal. pone. 0218594

Prabu M., Abhinav B. S., Narayanan B. H., Sugandhi S.S. and Rajeshkumar D. (2023). Design of water quality and soil macronutrients measuring device using TDS and NPK sensor. ARPN Journal of Engineering and Applied Sciences, 18(16):1923-1931

Qi H., Paz-Kagan T., Karnieli A., Li S.(2017). Linear Multi-Task Learning for Predicting Soil Properties Using Field Spectroscopy. Remote Sensing,

9(11):1099. https://doi.org/10.3390/rs9111099

Pattanayak U. and Mallick M. (2017). Agricultural Production and Economic Growth in India: An Econometric Analysis. Asian Journal of Multidisciplinary Studies, 5(3):61-66

Pamungkas, W., Afandi, M. A., Astiti, S., Enriko, I. (2023). Research in Electronic Multi-Sensor Accuracy in the Implementation of Soil Fertility Monitoring System Using LoRa. International Journal on Science, Engineering, Information technology, 13(6):2397-2406.

Sampath A. and Sumithira T. R. (2022). Sparse based recurrent neural network long shortterm memory (rnn-lstm) model for the classification of ECG signals. Applied Artificial Intelligence,36:1,1457-1488. "doi:10.1080/08839514. 2021. 2018183

Sampath A. and Sumithira T. R. (2022). Sparse based recurrent neural network long shortterm memory (rnn-lstm) model for the classification of ECG signals. Applied Artificial Intelligence, 36:1,1457-1488. doi:10.1080/08839514. 2021. 2018183

Shakoor A. and Ullah Z. (2024). Review of Agricultural Water Management Techniques for Drought Resilience and Water Conservation. International Journal of Research and Advances in Agricultural Sciences, 3(1): 31-45.

Simon, S.M., Glaum, P., Valdovinos, F.S. (2023). Interpreting random forest analysis of ecological models to move from prediction to explanation. Sci. Rep. 13,3881. https://doi.org/10.1038/s41598-023-30313-8

Somasundaram P. (2018). Efficient File-Based Data Ingestion for Cloud Analytics: A Framework for Extracting and Converting Non-Traditional Data Sources. International Journal of Science and Research, 13(2):2223-2227.

Subeesh A. and Mehta C.R.(2021). Automation and digitization of agriculture using artificial intelligence and internet of things. Artificial Intelligence in Agriculture 5:278–291.

TyralisH. and PapacharalampousG. (2017). Variable Selection in Time Series Forecasting Using Random Forests. Algorithms, 10(4):114.https://doi.org/10.3390/a10040114

van Houdt G., Mosquera C., Nápoles, G. (2020). A review on the longshort-term memory model. Artif Intell Rev 53,5929–5955. https://doi.org/10.1007/s10462-020-09838-1

Vinutha H.P., Poornima B., Sagar B.M. (2018). Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In: Satapathy, S., Tavares, J., Bhateja, V., Mohanty, J. (eds) Information and Decision Sciences. Advances in Intelligent Systems and Computing, Vol 701. Springer, Singapore. https://doi.org/10.1007/978-981-10-7563-6_53

Wang H.F.,Takematsu N., Ambe S. (2000). Effects of soil acidity on the uptake of trace elements in soybean and tomato plants. Appl Radiat Isot.,52(4):803-11. doi: 10.1016/s0969-8043(99)00153-0.

Wibawa, Whetton R.L., Zhao Y., Nawar S., Mouazen A.M. (2021). Modelling the Influence of Soil Properties on Crop Yields Using a Non-Linear NFIR Model and Laboratory Data. Soil Systems, 5(1):12. https://doi.org/10.3390/soilsystems 5010012