ML-Based Crop Water Requirement Prediction System

Dr.Thenmozhi R
Dept. of Computing Technologies
SRMIST
Chennai,India
thenmozr@srmist.edu.in

Makadia Yakshkumar Vijaykumar
Dept. of Computing Technologies
SRMIST
Chennai,India
mv1421@srmist.edu.in

Diva Alpeshkumar Merja
Dept. of Computing Technologies
SRMIST
Chennai,India
dm9283@srmist.edu.in

Abstract

The idea of using technology-driven precision agriculture has gained notable momentum in call to action to the ever growing global need for agriculture and freshwater [1]. With dwindling cultivable land and constrained water resources, particularly in semi-arid regions like Rajkot, Gujarat, there is an urgent need for intelligent irrigation systems guided by data analytics. While numerous machine learning models have been proposed for irrigation scheduling, many fall short in adapting to erratic climatic behavior—such as Gujarat's inconsistent monsoon patterns and temperature shifts—which affects their accuracy and usability.

To address this, the present study introduces an improved LSTM (Long Short-Term Memory) architecture tailored to Gujarat's specific agro-climatic context. The model offers short-term forecasting of soil moisture levels, optimal irrigation windows, and spatial water requirements for key crops such as wheat, cotton, and ground nuts. By leveraging historical climate, soil, and crop data, the system demonstrates improved irrigation efficiency in simulations, providing a scalable solution for farmers contending with limited water availability [2].

Keywords – Machine Learning, Water Prediction, LSTM, Agro-climatic Conditions

I. Introduction

The world population is estimated to sky rocket and touch aaprox. 10 billion by 2050, increasing annually by 1.1% from the current 7.7 billion. This rising population, combined with accelerated urbanization and growing influence of climatic conditions, puts immense pressure on agriculture land and freshwater availability [3]. Agriculture remains a foundational pillar for food security and economic development, yet only ~1% of Global water is easily available for irrigation.In which crop water requirement accounts for ~70% of world's fresh water , efficient water use is now more essential than ever [4].

This issue is particularly critical in semi-arid areas such as Gujarat, where erratic monsoon behavior and intensifying water scarcity pose serious risks to crop productivity. Crops like cotton, groundnuts, and wheat—key staples in Gujarat—require precise water management to sustain yield and conserve resources. Traditional irrigation methods still dominate, often relying on visual crop assessment, which results in up to 50% water wastage. In contrast, modern systems such as drip and sprinkler irrigation can reduce losses by 35–75%, yet their open nature often fails to contain soil moisture at optimal levels, leading to under- or over-irrigation [5]. These extremes not only impact yield but also elevate costs through excess energy use and nutrient leaching.

Precision irrigation technologies address these inefficiencies by using inputs like real-time soil moisture levels, local climatic weather, and crop and its specific requirements to fine-tune water delivery [6]. In regions with fluctuating rainfall and temperature patterns like Gujarat, such data influenced systems empowered and encouraged farmers to better adapt and shift to climate variability.

The use of ML in agriculture has reshaped traditional water management paradigms. Unlike mechanistic models—which require detailed calibration and often struggle with nonlinear, dynamic interactions in agro-systems—machine learning models analyse patterns from historical data, bypassing the need for complex assumptions. Among these, Long Short Term Memory (LSTM) networks is an expert at time/series prediction making, leveraging memory cells to capture long-term dependencies in data. Making them superior for modeling soil moisture patterns, which depend on a range of contributors like weather, crop growth, and irrigation events [7].

In Gujarat's agricultural landscape, LSTM models are particularly advantageous in capturing seasonal anomalies such as early heatwaves and delayed post-monsoon rains, thus offering precise irrigation insights. Recent advances stress the importance of high-resolution, localized datasets for improving prediction accuracy. For Gujarat, this includes variables like daily temperature shifts, humidity, wind speed, radiation levels, soil profile, and phenological crop stages.

While IoT-based systems enable real-time data collection, machine learning models based purely on historical and forecast data can offer comparable accuracy, making them more feasible for widespread adoption—especially among small and mid-sized farms with limited access to advanced tech infrastructure.

This paper presents an enhanced LSTM-based prediction framework, specifically designed for Gujarat's agricultural context. By relying solely on data-driven machine learning techniques—instead of infrastructure-heavy IoT systems—we offer a cost-effective, scalable, and climate-resilient solution.

The system integrates weather and soil inputs to forecast soil moisture content, determine irrigation timings, and optimize water distribution. Its closed-loop architecture allows real-time feedback updates, making it more responsive to climatic fluctuations than conventional models. This approach aligns with the region's pressing need for sustainable and efficient irrigation strategies tailored to its environmental challenges.

II. RELATED WORK

Recent studies have shown notable progress in applying ML techniques for irrigation management. [8] P. K. Kashyap et al. (2021) introduced an IoT-enabled intelligent irrigation system using deep learning neural networks to predict water needs based on real-time sensor data. Their system, while effective in improving irrigation efficiency, heavily relies on IoT infrastructure, which may not be easily scalable for smaller farms, especially in developing areas like Gujarat. This study differs by using LSTM networks that not only process real-time data but also integrate historical weather patterns, enabling better adaptability to varying climate conditions in semi-arid regions.

Similarly, J. Zhang et al. (2018) developed a deep learning model using LSTM networks to predict groundwater levels in agricultural areas. [9]Their model showed strong performance in capturing temporal water table variations, which is critical for optimizing irrigation in regions dependent on groundwater. While their research is focused on groundwater, our system expands on this by predicting soil moisture content in addition to groundwater levels, providing a more comprehensive solution for irrigation management in Gujarat's vibrant agriculture framework.

"Y. Park et al. (2009) proposed a receding horizon control algorithm for adaptive management of soil moisture and chemical levels during irrigation." [10] While effective, this method uses a fixed algorithm that may not fully adapt to unpredictable climatic changes, unlike the LSTM-based model used in this study, which learns from historical and real-time data, making it more responsive to fluctuating environmental conditions.

"Kamilaris and F. X. Prenafeta-Boldú (2018) surveyed deep learning applications in agriculture, focusing on general tasks like yield prediction and disease detection." [11] However, their review doesn't address the specific challenges of irrigation in semi-arid regions. Our approach, leveraging LSTM networks, predicts water needs by incorporating climatic and soil data, offering a more tailored solution for water-scarce regions like Gujarat.

In [12] J. Liakos et al.'s (2018) review, the authors discussed machine learning applications in agriculture, emphasizing crop prediction but not specifically addressing water management. Our study goes further by integrating localized climate and soil data into the irrigation process, making it more applicable for regions with specific water challenges, such as Gujarat.

"N. Efremova et al. (2019) developed a model for soil moisture prediction using satellite data and sequence-to-sequence networks." [13] While satellite data can be useful, our system integrates local soil and climatic factors, offering more accurate and feasible solutions for small-scale farmers in rural areas where satellite data may not be as detailed.

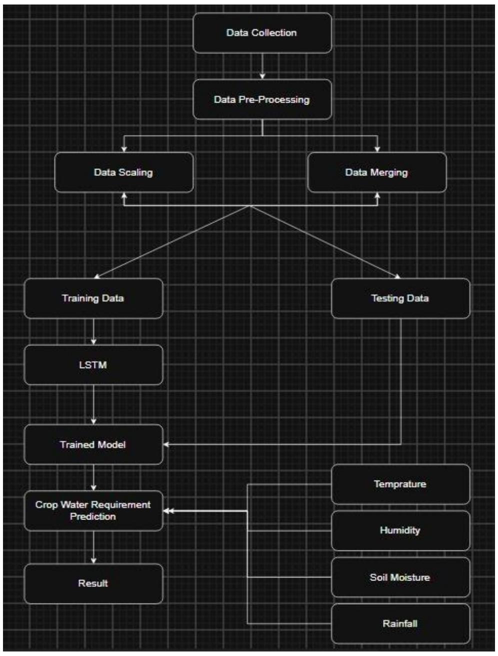
"Y. Saikai et al. (2023) employed deep reinforcement learning for irrigation scheduling using sensor feedback. "[14] However, this method requires extensive sensor networks, making it less accessible for small farms. In contrast, our system uses historical weather and soil data, which is less infrastructure-intensive and more feasible for farmers in regions like Gujarat with limited resources.

Collectively, these studies underscore the importance of machine learning in irrigation management but also reveal gaps, especially in addressing the unique needs of semi-arid regions. Our approach, which uses LSTM networks to predict both soil moisture and water table depth, provides a more scalable and adaptable solution to the specific irrigation challenges faced by Gujarat's farmers.

III. TRAINING PROCEDUCE

To accurately predict volumetric soil moisture and improve irrigation scheduling, a data-driven approach was adopted using Long Short Term Memory (LSTM) networks. The training procedure followed a systematic pipeline comprising four major stages: dataset collection, data preprocessing, model training, and model testing. This section outlines each stage in detail, emphasizing the sources, preparation, and modeling strategy tailored for the agricultural context of Rajkot, Gujarat.

A. *Dataset Collection*



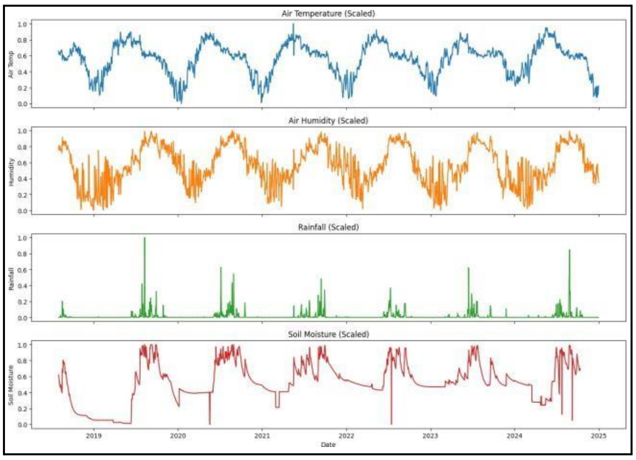
To ensure accurate soil moisture prediction, multiple datasets encompassing climate, crop physiology, soil characteristics, and historical moisture records were gathered. This comprehensive integration supports time-based learning and enhances agronomic relevance, especially for the Rajkot district.

The foundation of the model was established using climate data retrieved from the NASA POWER database, focused on Rajkot's coordinates. Spanning from 2018 onwards, this dataset provides essential meteorological variables that influence evapotranspiration and soil moisture levels. These include daily records of {"air temperature (mean, maximum, and minimum), relative humidity, precipitation, wind speed, and solar radiation"}. Together, these inputs capture the environmental dynamics impacting irrigation demand and natural water availability.

To model crop-specific water usage, a crop calendar was constructed for the five major crops grown in the region: Wheat, Groundnut, Cotton, Chickpea, and Cumin. Information was compiled from government agricultural resources, detailing growth stage durations, root depths, and crop coefficients. These elements are essential in calculating daily crop water requirements and capturing how moisture needs vary throughout a crop's life cycle. The inclusion of such physiological parameters allows the model to simulate real-world agricultural practices with greater precision.

Soil characteristics were captured through a soil classification dataset developed from regional agricultural and environmental data. The dataset included a detailed profile of soil up to 200 cm depth, with a specific focus on the 0–15 cm layer relevant for moisture prediction. Key features included the percentage composition of sand and clay, bulk density (linked to the soil's water holding capability), and organic carbon amount (an indicator of soil fertility). Although these attributes do not change over time, they provide a necessary baseline to understand how different soil types influence water absorption and storage.

The target variable—volumetric soil's moisture at 15 cm depth—was collected from the NRSC's Variable Infiltration Capacity (VIC) model, which offers region-specific soil moisture estimates. These daily measurements were critical for training the machine learning model. To build a cohesive training dataset, temporal alignment was performed between the moisture values and the corresponding climate and crop data. This ensured consistency across input variables and improved the reliability of the model's learning process.



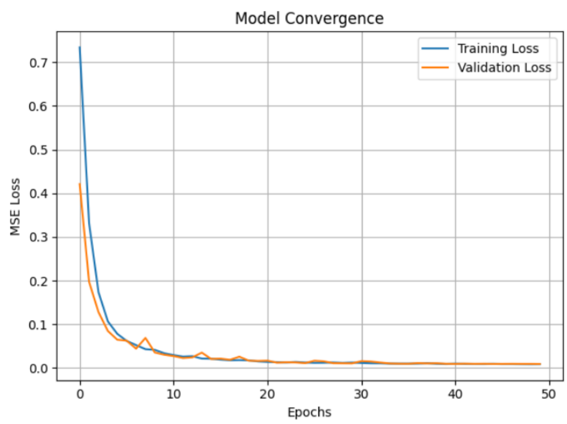
B. Data Preprocessing and Model Training

Prior to model training, a comprehensive preprocessing pipeline was implemented to structure and sanitize the raw input data for machine learning applications. All relevant datasets—including climate variables, crop data, and soil attributes—were merged based on the date to form a single, unified dataset. This consolidation ensured temporal alignment across all feature domains. Records containing missing values, particularly those lacking soil moisture measurements (the prediction target), were removed to maintain data integrity. Following this, numeric values were normalized using "Min-Max scaling" to constrain their values between 0 and 1. This step was critical for ensuring feature uniformity and improving the learning efficiency of the Long Short Term Memory (LSTM) model.

$$L(W^t) = \left[\frac{1}{r}\sum_{t=1}^{r}(y_{obs} - y_{pre})^2\right]^{0.5}$$

To convert the time-series data into a supervised learning standard , a sliding window approach was employed. In this method, sequences were generated using a fixed window of historical data—for example, a 7-day span—with the soil moisture of the following serving as the prediction target. These input-output pairs were then reshaped into three-dimensional tensors, the format required by LSTM models, to preserve the temporal structure and feature relationships across time steps.

The predictive model itself was designed using a multi-layer LSTM architecture, chosen for its ability to capture long range dependencies in sequential and historic data. Through its internal memory and gating mechanisms, the LSTM network learned to model complex interactions between climatic conditions, soil properties, and crop-related factors over time. The database was partitioned into testing and training subsets using a 70:30 spliting to evaluate performance reliably.



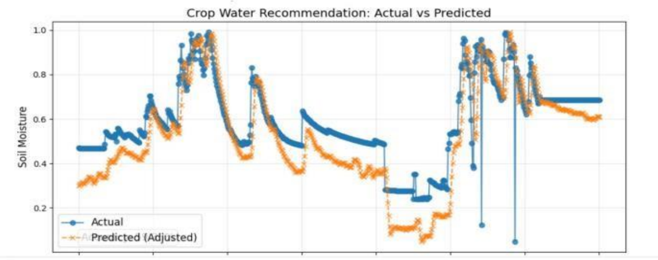
To optimize learning, training was conducted using mini-batches, with batch sizes selected as powers of two (e.g., 32 or 64) to improve memory management and convergence behavior. The model aimed to minimize prediction error by optimizing for the Root Mean Squared Error (RMSE) between actual and predicted soil moisture numerics.

Several techniques were employed to ensure effective training and model generalization.We used early stopping to automatically end training when the model stopped improving on validation data, which helped avoid overfitting. Additionally, dynamic learning rate scheduling was used to adjust the optimization step size based on training progress, enhancing both training speed and accuracy. These strategies collectively ensured a robust and efficient convergence of the predictive model.

C. Model Testing

After training the model, we tested its performance on a separate dataset we'd held back for this purpose. We used two common accuracy measures - MAE and RMSE - to see how well it predicted soil moisture levels. With an MAE score of 0.2008 and RMSE of 0.2983, the model proved quite accurate, showing only small differences between its predictions and the actual moisture measurements.

These results confirmed the model's validity to learn and generalize patterns in the soil moisture data effectively. Visual inspection of the predicted time series further supported this, as the model outputs aligned well with the actual moisture trends observed in the test data. This strong predictive performance underscores the model's potential to serve as a core component in intelligent irrigation systems. By leveraging forecasted soil moisture and weather conditions, the system can enable proactive, data-driven irrigation scheduling that improves resource efficiency and crop health.



D. Formulae and Calculations

To create predictive features that reflect the soil–plant–atmosphere continuum, several agronomic and meteorological equations were applied. These transformations were essential for converting raw datasets into usable inputs for the LSTM model.

1. "Crop Water Requirement (ETc)"

The actual crop water requirement, also known as evapotranspiration (ETc), was computed using the crop coefficient method.

$$ET_c = K_c \times ET_o$$

Where:

- The crop coefficient (Kc) changes depending on what crop you're growing and its current growth phase.
- Reference evapotranspiration (ETo) can come from weather databases or be calculated using local climate data.

2. Reference Evapotranspiration (ETo)

When ETo was not directly available, it was estimated using the Hargreaves method, which calculates reference evapotranspiration based on daily temperature extremes and solar radiation.

$$ET_o = 0.0023 \times (T_{avg} + 17.8) \times (T_{max} - T_{min})^{0.5} \times Ra$$

Where:

- **Tavg, Tmax, Tmin** are the average, maximum, and minimum daily temperatures (°C).
- **Ra** is extraterrestrial radiation, derived using geographic coordinates and the Julian day.

3. Effective Rainfall (Peff)

Effective rainfall refers to the proportion of precipitation usable by crops after accounting for losses. It was estimated using empirical relationships.

$$P_{eff} = P_{total} \times 0.8 \quad \text{(if daily rainfall is moderate)}$$

Where:

- **Ptotal** represents the total daily precipitation.

4. Water Holding Capacity (WHC)

Soil water retention was modeled using the water holding capacity, calculated based on soil bulk density and effective rooting depth.

$$WHC = d \times BD \times \theta$$

Where:

- **d** is the root zone depth (in meters),
- **BD** is the soil bulk density.
- **θ** is the available water content determined from soil texture.

To quantify model performance during training and testing, the following evaluation metrics were used:

- **"Root Mean Squared Error (RMSE)"**

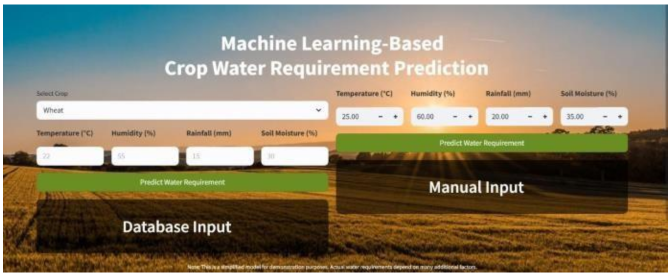
$$L(W^t) = \left[\frac{1}{r}\sum_{t=1}^{r}(y_{obs} - y_{pre})^2\right]^{0.5}$$

- **"Mean Absolute Error (MAE)"**

These metrics captured both large and average deviations between actual and predicted soil moisture values, helping fine-tune model performance

E. *Website Deployment*

The target variable—volumetric soil moisture at 15 cm depth—was collected from the NRSC's Variable Infiltration Capacity (VIC) model, which offers region-specific soil moisture estimates. These daily measurements were critical for training the machine learning model. To build a cohesive training dataset, temporal alignment was performed between the moisture values and the corresponding climate and crop data. This ensured consistency across input variables and improved the reliability of the model's learning process.



IV. RESULT AND CONCLUSION

To ensure broad usability, the trained LSTM model was deployed through a user-friendly website interface, allowing farmers and agricultural planners to interact with the system without requiring technical knowledge.

The model's performance was evaluated using two standard regression metrics — **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** — to assess its accuracy in predicting daily volumetric soil moisture. The results on the test dataset were as follows:

- **Mean Absolute Error (MAE):** 0.2008
- **Root Mean Squared Error (RMSE):** 0.2983

These values indicate that the model reliably captures the dynamics of soil moisture influenced by climatic and agronomic factors. The low MAE reflects high average accuracy in predictions, while the RMSE further confirms the model's robustness under varying environmental conditions.

A comparative plot of actual versus predicted soil moisture illustrates a close alignment between the two time series, especially during periods of low to moderate fluctuations, highlighting the LSTM's ability to model temporal dependencies.

The preprocessing and feature engineering phase was vital to the model's success. By transforming raw environmental and crop data into structured input variables — such as ETc, effective rainfall, and soil water holding capacity — the system gained a contextual understanding of the soil-water-plant relationship.

This project integrates the following components to form a practical and scalable intelligent irrigation framework:

- **Climate data** obtained from NASA POWER
- **Crop calendars** specific to regional crops like wheat, groundnut, cotton, chickpea, and cumin
- **Soil classification** features tailored to Rajkot's local geography
- **LSTM-based forecasting** for capturing temporal soil moisture trends

Beyond achieving solid performance on historical data, the solution establishes a scalable pathway for intelligent irrigation management. With future enhancements — such as real-time sensor integration and support for more diverse crop types — the system has the potential to deliver even more precise recommendations and broader applicability across regions.

By bridging advanced machine learning with grassroots-level farming needs through an intuitive interface, this work presents a step forward in sustainable and data-driven agriculture.

REFERENCES

[1] Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. Computers and Electronics in Agriculture, 36(2–3), 113–132. https://doi.org/10.1016/S0168-1699(02)00096-0

[2] Shahid, M., Mistry, D., & Gupta, H. (2021). Crop Water Requirement Prediction Using Deep Learning Approaches. Journal of Water and Climate Change, 12(3), 734–747. https://doi.org/10.2166/wcc.2021.144

[3] United Nations, Department of Economic and Social Affairs, Population Division. (2019). World Population Prospects 2019: Highlights. https://population.un.org/wpp/

[4] FAO. (2011). The State of the World's Land and Water Resources for Food and Agriculture (SOLAW). Food and Agriculture Organization of the United Nations. https://www.fao.org/3/i1688e/i1688e.pdf

[5] Perry, C., Steduto, P., Allen, R. G., & Burt, C. M. (2009). Increasing productivity in irrigated agriculture: Agronomic constraints and hydrological realities. Agricultural Water Management, 96(11), 1517–1524.

[6] Jones, H. G. (2004). Irrigation scheduling: Advantages and pitfalls of plant-based methods. Journal of Experimental Botany, 55(407), 2427–2436.

[7] Zhao, Y., Zhang, N., & Wang, G. (2020). Soil moisture prediction using LSTM networks with climatic and irrigation data. Computers and Electronics in Agriculture, 173, 105393. https://doi.org/10.1016/j.compag.2020.105393

[8] Kashyap, P. K., Kumar, S., Jaiswal, A., Prasad, M., & Gandomi, A. H. (2021). Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network. IEEE Internet of Things Journal, 8(14), 11345-11358.

[9] Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. Journal of Hydrology, 561, 918-929.

[10] Park, Y., Shamma, J. S., & Harmon, T. C. (2009). A receding horizon control algorithm for adaptive management

of soil moisture and chemical levels during irrigation. Environmental Modelling & Software, 24(7), 928–940.

[11] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90.

[12] Liakos, K. G., et al. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[13] Efremova, N., Zausaev, D., & Antipov, G. (2019). Prediction of soil moisture content based on satellite data and sequence-to-sequence networks. arXiv preprint arXiv:1906.01484.

[14] Saikai, Y., Peake, A., & Chenu, K. (2023). Deep reinforcement learning for irrigation scheduling using high-dimensional sensor feedback. arXiv preprint arXiv:2301.00543.

[15] Mortazavizadeh, F., Bolonio, D., Mirzaei, M., Ng, J. L., Mortazavizadeh, S. V., Dehghani, A., Mortezavi, S., & Ghadirzadeh, H. (2025). Advances in machine learning for agricultural water management: A review of techniques and applications. Journal of Hydroinformatics, 27(3), 474-492.

[16] Sharma, A., Shukla, A. K., Rao, K. H., Singh, M., Muniyandy, E., & Sridhar, S. (2023). Machine learning and sensor-cloud based precision agriculture for intelligent water management for enhanced crop productivity. *Journal of Environmental Science*, 12(3), 45-52.

[17] Buakum, B., Limpinuntana, V., Vorasoot, N., Pannangpetch, K., & Bell, R. W. (2011). Rooting patterns of four crop legumes in response to seed-placement depths in the dry season. Plant Production Science, 14(1), 35-48.

[18] Sharma, S., Rai, S., & Krishnan, N. C. (2020). Wheat crop yield prediction using deep LSTM model. arXiv preprint arXiv:2011.01498.

[19] Reddy, L. V., Ganesh, D., Kumar, M. S., Gogula, S., Rekha, M., & Sehgal, A. (2024). Applying machine learning to soil analysis for accurate farming. Journal of Agricultural Technology, 20(1), 1-15.

[20] Peng, Y., He, M., Zheng, Z., & He, Y. (2023). Enhanced neural network for rapid identification of crop water and nitrogen content using multispectral imaging. Agronomy, 13(10), 2464.

[21] Munir, M. S., Bajwa, I. S., Naeem, M. A., & Ramzan, B. (2018). Design and implementation of an IoT system for smart energy consumption and smart irrigation in tunnel farming. Energies, 11(12), 3427.

[22] Wang, Y., Velswamy, K., & Huang, B. (2017). A Long-Short Term Memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. Processes, 5(3), 46.

[23] Delgoda, D., Saleem, S. K., Malano, H., & Halgamuge, M. N. (2015). Root zone soil moisture prediction models based on system identification: Formulation of the theory and validation using field and AQUACROP data. Agricultural Water Management, 163, 11-31.

[24] Bellido-Jiménez, J. A., Estévez, J., & García-Marín, A. P. (2021). New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. Agricultural and Forest Meteorology, 297, 108242.

[25] Liu, D., Yu, Z., & L, H. (2015). Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. Water Resources Research, 51(4), 2242-2261.