International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparison of AI Techniques for soil moisture prediction

Shreyas Dhane[1], Dr. Aarti Karande[2], Pritish Dutta[3], Niketan Gawade[4]

[1]*Master of Computer Applications Sardar Patel Institute of Technology* Mumbai, India shreyas.dhane23@spit.ac.in
[2]*Master of Computer Applications Sardar Patel Institute of Technology* Mumbai, India aartimkarande@spit.ac.in
[3]*Master of Computer Applications Sardar Patel Institute of Technology* Mumbai, India pritish.dutta23@spit.ac.in
[4]*Master of Computer Applications Sardar Patel Institute of Technology* Mumbai, India niketan.gawade23@spit.ac.in

Abstract—

In our paper, we explored how Machine Learning (ML) can help in managing irrigation and using water more efficiently in agriculture. Water is a very important resource, and managing it well is crucial for both farming and other industries. We explored different ML techniques to solve common problems like overuse of water, underuse, and unpredictable weather. We used models like Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks to see how ML can automate irrigation schedules, predict soil moisture, and adjust to changing weather conditions. The main aim was to reduce water wastage and improve crop yields.

What we learned is that these models, especially LSTM, are really helpful in analyzing data related to weather, soil, and crops. By using past data, these models help in controlling irrigation more precisely, meaning farmers can water their crops only when needed and in the right amount, which saves water. We also thought about how ML could be used outside of farming, like in industries and cities, to predict water demand, monitor water quality, and find inefficiencies. The biggest finding from our research was that LSTM net- works are the best for irrigation management. While SVR, KNN, and ANN are useful, LSTMs are particularly good at handling time-series data, which is very important for irrigation schedules and predicting soil moisture. This makes LSTM the most reliable choice for managing irrigation in a way that adapts to changes in weather and soil conditions over time.

I. Introduction

Managing water resources effectively is a critical issue, especially in agriculture, industry, and urban areas. Inefficient water use can harm crops, while industries not only consume large amounts of water but also contribute to pollution. In countries like India, the rising water demands of industries, coupled with improper disposal of wastewater, are making the situation worse.

On top of that, cities often face droughts and water short- ages, making it hard for officials to ensure a steady water supply for everyone. Machine learning (ML) is transforming how farmers manage water for their crops, especially when it comes to understanding soil moisture and planning irrigation.

Identify applicable funding agency here. If none, delete this.

By analyzing past data like weather patterns, soil types, and crop needs, ML models can predict how much water the soil will retain at different times. This helps farmers know exactly when and how much to water, avoiding overwatering or underwatering and ensuring crops get just the right amount of moisture.

ML also takes the guesswork out of irrigation. Using smart algorithms, it creates irrigation schedules that adapt to changing conditions, like unexpected rainfall or heatwaves. It can even divide fields into zones based on water needs, making irrigation more precise and efficient. With ML, farmers can save water, cut costs, and boost crop yields—all while being kinder to the environment.

II. Technology and challenges

A. Crop Irrigation Technology and Challenges

Irrigation is a delicate process where too much or too little water can damage crops. Traditional methods, where humans control irrigation, are not ideal for large farms or when quick action is needed, especially with unpredictable weather and potential animal interference. This is where intelligent systems like machine learning, and deep learning can help. By automating the process and using data from soil moisture sensors and weather forecasts, farmers can ensure water is used efficiently, cutting down on labor and waste.

III. Automation with Machine Learning

Machine learning plays a crucial role in modern irrigation systems by analyzing data to optimize water usage.[1] Models like LSTM and Artificial Neural Networks (ANNs) are used to predict soil moisture levels, assess drought risks, and recommend optimal irrigation schedules. These algorithms analyze historical weather data, crop types, and soil characteristics to ensure precise water delivery, minimizing waste and improving crop productivity. Over time, these models learn and adapt to changing environmental conditions, making the system more efficient and reliable.

Beyond agriculture, ML has vast potential in industrial and urban water management. It can optimize water distribution by analyzing consumption patterns, identifying inefficiencies, and forecasting future water demands. Additionally, ML models can monitor water quality in real time by detecting anomalies in data, such as sudden contamination or changes in chemical levels. By integrating predictive analytics and real-time monitoring, ML empowers industries and cities to manage water resources more sustainably and effectively.

A. Abbreviations and Acronyms

Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized various domains, including agriculture and irrigation management.[1] Leveraging these technologies enable efficient and automated systems for monitoring and decision-making. [5] Among the ML models, Artificial Neural Networks (ANN), D Support Vector Regression (SVR), and K- Nearest Neighbors (KNN) are widely employed for predictive analytics and classification tasks. Advanced techniques like Long Short-Term Memory (LSTM) networks and Convolu- tional Neural Networks (CNN) are particularly effective in handling time-series data and spatial imagery, respectively. Performance evaluation of these models is often conducted using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), ensuring robust and accurate solutions for irrigation optimization and resource management.

B. Units

This paper primarily adopts the SI (MKS) unit system for measurements and analysis. For ease of understanding and global standardization, the following conventions were observed: All calculations and results are presented in SI units. Any alternative units, if required, are included in parentheses for reference. For example, temperature is recorded in degrees Celsius (°C). Consistent unit rep- resentation is maintained throughout the paper, such as ”webers per square meter (Wb/m^2)” instead of a mixture like ”webers/m^2.” Units are spelled out in textual expla- nations when needed (e.g., ”a few henries” rather than ”a few H”). Decimal notation strictly includes a leading zero (e.g., ”0.25” instead of ”.25”). For parameters such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), the respective units depend on the quantity being evaluated and are explicitly stated alongside the results.

C. Justification

As water becomes scarcer and the impacts of climate change grow more severe, we need smarter ways to manage it.[1] Automating irrigation and using machine learning addresses problems like water waste, manual labor, and unpredictable weather, making it easier to ensure crops and industries get just the right amount of water. These systems not only improve farming efficiency but also help industries reduce their water footprint and manage pollution more responsibly.

D. Challenges solved earlier

- 1 Moisture Temperature Hydrological processes are highly non-linear, dynamic, and exhibit variability across temporal and spatial scales, making them difficult to model accurately with traditional linear statistical meth- ods. Solution: Advanced AI techniques such as Artificial Neural Networks (ANNs)
- 2.[4] Traditional linear models often require extensive explanatory variables to capture the complex physical processes in hydrology. This can limit their applicability, especially when high-quality data are scarce. Solution: Data-driven AI models, SVMs
- 3. Changing climate conditions and anthropogenic fac- tors are altering hydrological patterns, making traditional statistical models less effective. Solution: Hybrid models, such as wavelet-transformed neural networks
- 4. Water Quality Prediction and Classification Model: Random Forest (RF) Predicting water quality based on multiple parameters (e.g., pH, turbidity, contaminants). Solution: RF models can handle large feature sets effectively, providing robust classification or regression based on historical data. By training on water quality datasets, RF can predict levels of contaminants or categorize water quality into different safety levels, helping authorities take preventive measures.
- 5 . Irrigation Control Problem: Over-extraction of ground- water poses a threat to water availability for other purposes, necessitating precise irrigation control. Solutions: Deep Learning Classifiers: Implementing various deep learning architectures (AlexNet, GoogleNet, ResNet, VGG16, SqueezeNet) to classify soil texture and irrigation needs effectively. AlexNet achieved the highest performance with an F1 score of 0.9973.[3] UAV Velocity Measurement: Using drones and the YOLOv5 algorithm to measure surface flow velocity in rivers, enhancing decision-making in irrigation management.
- 6. Water Level Forecasting Model: Long Short-Term Memory (LSTM) Networks Time-series prediction for river or reservoir water levels. Solution: LSTM models capture temporal dependencies, making them ideal for forecasting based on historical water level data. [7]They help in anticipating fluctuations in water levels, sup- porting decision-making for reservoir management, and optimizing water storage.

- 7. Data Scarcity and Quality Problem: Insufficient or low- quality labeled data can lead to inaccurate classification results, affecting decision-making in water resource management. Solution: Implement data augmentation techniques to artificially increase the size and variability of the training dataset. [2]Collaborate with local communities and organizations to gather high-quality, labeled data. Utilize transfer learning from pre-trained models to leverage existing knowledge and improve classification accuracy.
- 8. Prediction Limitations Problem: Machine learning models, such as LSTM for flood forecasting, demonstrate high accuracy for short-term predictions but struggle with peak value predictions over longer lead times. Solution: Enhance model performance by integrating additional data sources, such as historical flood data and climate models. Use ensemble methods that combine predictions from multiple models to improve accuracy over various lead times. Regularly update and retrain models with new data to adapt to changing conditions.
- 9. Environmental Threats to Water Resources Problem: Pollution, climate change, and over-extraction of ground- water threaten water resources, complicating management efforts. Solution: Develop comprehensive monitoring systems using real time data. Implement strict regulations on groundwater extraction and pollution control measures. Promote sustainable practices such as rainwater harvesting and water recycling to reduce dependency on existing water sources.
- 10. Limitations of Rainfall Forecast Models Problem: Traditional radar-based precipitation forecasts can be un- reliable during heavy rainfall, indicating inadequacies in current forecasting methods. Solution: Invest in advanced machine learning techniques, such as deep learning mod- els that can better analyze complex weather patterns. Utilize hybrid models that combine radar data with ground- based observations and satellite imagery for improved accuracy. Conduct ongoing research to refine forecasting models and adapt them to local conditions.
- 11. Inefficiency of Existing Classification Techniques Problem: Different classification methods yield varying effectiveness in land cover change detection post-flood, which may lead to mismanagement of land and water resources. Solution: Explore and implement hybrid classification methods that combine the strengths of multiple algorithms (e.g., merging supervised and unsupervised approaches). Invest in the development of advanced algorithms like deep learning-based image segmentation techniques, which can improve classification accuracy. Regularly validate and calibrate classification models against ground truth data to ensure reliability.
- 12. Rainwater Management Problem: Rainfall prediction is crucial for effective water harvesting and management, especially with irregular datasets like rainfall. Solutions: Hybrid Models: Combination of machine learning (ML) and deep learning (DL) models (e.g., CNN, SVM, Linear Regression) demonstrates higher accuracy in predicting rainfall compared to traditional ML methods. Automated Solutions: Utilizing computer vision and deep learning to automate the assessment of rooftop catchment volumes and optimal tank placements, ensuring efficient rainwater harvesting.
- 13: Predicting Water Quality Under Changing Environ- mental Conditions Water quality can fluctuate significantly due to changing environmental conditions, such as rainfall, temperature, and land usage. Predicting these changes accurately is crucial for timely interventions.

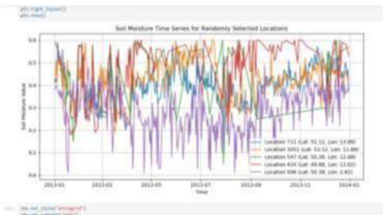


Fig. 1. Soil moisture data

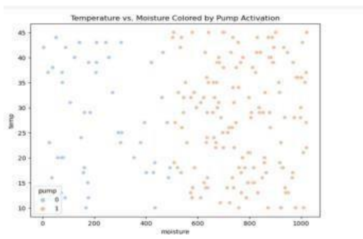


Fig. 3. Temperature vs Moisture colored by Pump Activation

Solution: Dynamic Machine Learning Models for Time- Series Forecasting Implement dynamic machine learning models that can adapt to real-time changes in environ- mental conditions. Time-series analysis techniques, such as Long Short-Term Memory (LSTM) networks, can be used to model the temporal dependencies in water quality data. By training these models on historical data that includes environmental factors, they can forecast future water quality trends under various scenarios. This proactive approach allows water management authorities to prepare for potential quality issues before they arise, enabling timely and effective responses.

E. Unresolved issues

Demand Forecasting and Resource Management Model: Time Series Models (ARIMA, Prophet) Predicting water demand based on historical data, population growth, and weather patterns. Solution: Time series models can forecast demand trends, helping water utilities to plan for infrastructure expansion and allocate resources efficiently. These predictions are useful for developing long-term strategies for water supply in high-demand regions.

F. LSTM (Long Short Term Memory

Preparing data : Handle Missing Data: Fill missing values using techniques like mean, median, or interpolation for numerical data. Feature Engineering: Normalize/Scale: Scale values to a range (e.g., 0 to 1) for better LSTM performance. Add Time Features: Include derived features like hour, day, or season for time-contextual patterns. Preprocessing data : Once the data is collected, it needs to be preprocessed to ensure

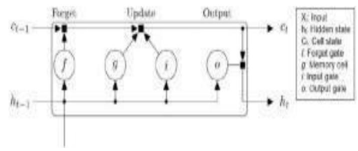


Fig. 4. LSTM

Fig. 5. Validation loss

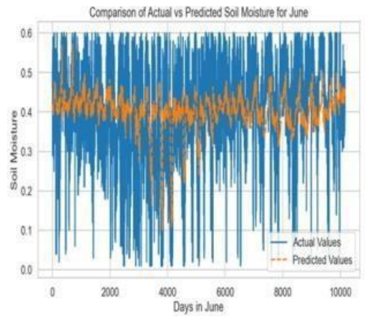


Fig. 6. Comparison of actual and predicted

it is in a suitable format for training the model. Handling Missing Data • Impute missing values using mean, median, or interpolation techniques if there are gaps in the data. Feature Engineering • Normalize or scale the data: LSTMs perform better when the data is normalized (scaled to a specific range like 0 to 1). • Time-based features: You may want to include time-based features like hour of the day, day of the week, or season as inputs.

G. ANN (Artificial Neural Networks)

Data Collection:

Collect data from various sources like soil moisture levels, weather forecasts, temperature, humidity, and crop types. Data Preprocessing:

Handle Missing Data: Impute missing values (mean, median, or interpolation). Feature Engineering: Normalize the data (scale values to a range of 0 to 1). Add time-based features (e.g., day of the week, season). Encode categorical data (e.g., crop type) if necessary. Splitting Data:

Split the dataset into training, validation, and test sets Model Construction:

Define the ANN architecture: Input layer: Features like soil moisture, temperature, etc. Hidden layers: A few dense layers with activation functions like ReLU. Output layer: Predicted irrigation amount or system state (e.g., irrigation on/off). Model Training:

Train the model using a suitable optimizer (e.g., Adam) and loss function (e.g., Mean Squared Error for regression tasks). Monitor performance using validation data to prevent overfitting. Model Evaluation:

Evaluate the model on the test set to check its accuracy or performance. Fine-tune hyperparameters (e.g., learning rate, number of hidden layers) based on evaluation results.

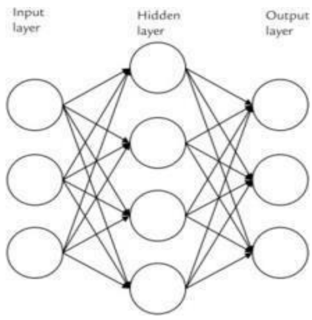


Fig. 7. ANN

```
[14]: loss = model.evaluate(X_test, y_test)
      print(f"Test Loss: {loss}")

      2010/2010 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - loss: 0.0059
      Test Loss: 0.00591687299311161
```

Fig. 8. Test loss

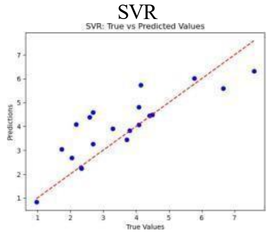
H. SVR (Support Vector Regression)

Data Collection:
Collect data such as soil moisture, temperature, humidity, weather forecasts, and crop type. Data Preprocessing:
Handle Missing Data: Impute missing values (mean, median, or interpolation). Feature Engineering: Normalize/scale data (using techniques like Min-Max scaling). Add time-based features (e.g., hour of the day, season).

Splitting Data:
Split the dataset into training, validation, and test sets Model Construction:
Define and train an SVR model using a kernel function (linear, polynomial, or RBF). Use Radial Basis Function (RBF) kernel for non-linear regression in most cases.

Model Training:
Train the model using the training data. Use C (regularization parameter) and epsilon (tolerance for error) for optimal results. Model Evaluation:
Evaluate the model on the test set to assess performance (using metrics like Mean Squared Error or R^2 score). Tune the hyperparameters (C, epsilon) to improve the model.

SVR



I. K-Nearest Neighbour (KNN)

Data Collection:
Gather data including soil moisture, temperature, rainfall, humidity, crop types, and historical irrigation data. Data Pre- processing:
Handle Missing Data: Impute missing values (mean, median, or interpolation). Feature Engineering: Normalize/scale the data (using Min-Max or Standard Scaling). Add time-

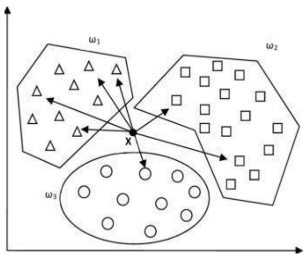


Fig. 10. KNN

```
Regression - Mean Squared Error: 0.01
Regression - R-squared: 0.58
```

Fig. 11. Regression

based features (e.g., time of day, season, or day of the week).
Splitting Data:
Split the dataset into training, validation, and test sets.

Model Construction:
Set the value of K (the number of neighbors) for the KNN algorithm. Choose an appropriate distance metric (e.g., Euclidean distance). Model Training:
Train the KNN model using the training data. The algorithm will learn by storing all the data points and making predictions based on the majority class or average of the K nearest neighbors. Model Evaluation:
Evaluate performance using test data and metrics such as accuracy, Mean Squared Error (MSE), or R^2 score. Tune the K value for optimal results.

Models	Mean Absolute Error	Mean Square Error	RMSE	R2 Score
ANN- Artificial Neural Networks	0.05807650815943019	0.005916872120282498	0.07692120722065209	0.5158737328999436
LSTM	0.08379808477147882	0.10339752118460263	0.3215548494185753	-1.6197744585408538
SVR	0.7120140825929	0.963052554648684	0.9813524110372808	0.636687140007224
K-Nearest Neighbour (KNN)	0.05205895797378609	0.005146446096086575	0.07173873437417654	0.5789110116267202

Comparison of models

IV. Conclusion

Long Short-Term Memory (LSTM) networks have played in advancing the field of irrigation management, particularly for predicting soil moisture levels. LSTM, with its unique ability to capture long-term dependencies in time-series data, has proven to be an ideal model for analyzing the temporal patterns inherent in soil moisture. This ability allows for more accurate predictions, which is critical for managing irrigation effectively and efficiently. By using LSTM, we've been able to optimize irrigation schedules, reducing water waste and ensuring crops receive the right amount of water at the right time.

The application of LSTM in irrigation management goes beyond just improving crop yield; it also contributes to the sustainable use of water resources, an issue that is becoming more urgent as the global population grows and climate change affects water availability.[6] LSTM's ability to forecast moisture levels based on historical data and weather conditions means that irrigation systems can adapt quickly to changes, helping farmers make better, data-driven decisions. This approach is not only beneficial for agriculture but also for the environment, as it helps minimize the overuse of water.

We also recognize the broader potential of LSTM in other domains of water management, where it could help optimize resource usage, predict future water availability, and even assist in areas such as water quality monitoring and wastewater management. As the world continues to face challenges related to water scarcity and resource conservation, the role of ma- chine learning, and specifically LSTM, becomes increasingly important in addressing these global issues.

ACKNOWLEDGEMENT

The research paper was supported by Sardar Patel Institute of Technology. Special thanks to Dr. Aarti Karande who provided insights and guidance that made a huge difference to our research.

REFERENCES:

[1] Atiya, A. F., El-Shoura, S., Shaheen, S. (1999). A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Transactions on Neural Networks, 10(2), 402-409. doi:10.1109/72.750569

[2] Chang, F.-J., Chang, Y.-T. (2006). Adaptive neuro- fuzzy inference system for prediction of water level in reservoir. Advances in Water Resources, 29(1), 1-10. doi:https://doi.org/10.1016/j.advwatres.2005.04.015

[3] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning , 20, 273-297. doi:https://doi.org/10.1007/BF00994018

[4] He, Z., Wen, X., Liu, H., Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference sys- tem and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology, 509, 379-386. doi:https://doi.org/10.1016/j.jhydrol.2013.11.054

[5] Shanker, M., Hu, M., Hung, M. (1996). Effect of data stan- dardization on neural network training. Omega, 24(4), 385-397. doi:https://doi.org/10.1016/0305-0483(96)00010-2

[6] Seo, Y., Kim, S., Kisi, O., Singh, V. (2015). Daily water level forecasting using wavelet decompositionand artificial intelligence

techniques. Journal of Hydrology, 520, 224-243. doi:https://doi.org/10.1016/j.jhydrol.2014.11.050

[7] Deployment and use of Artificial Intelligence (AI) in water resources and water management by Sakti Prajna Mahardika

[8] Parameswaran, G. and Sivaprasath, K. (2016) Arduino Based Smart Drip Irrigation

[9] Prasanna, G., Parvatham, S. and Krishna, S. (2016) Web Based Automatic Irrigation and Magazine of Engineering, Technology, Management and Research

[10] Tyagi, A., Gupta, N., Navani, J.P., Tiwari, M.R. and Gupta, M.A. (2017) Smart Irrigation System. International Journal for Innovative Research in Science & Technology, 3, 9-12.

[11] Smart Irrigation System. International Journal of Scientific Research in Science and Technology, 2, 343-345.