Soil Moisture Prediction using Integrated Supervised and Unsupervised Machine Learning Algorithms: Application to Precision Agriculture

CP303 - CAPSTONE PROJECT II

Supervisor:
Dr. Jayaram Valluru

Presented by:
Rahul Kumar Saw (2021CHB1052)

Upto Midsem: Recap

Model Discretization: Richard's Equation

↓

Open Loop Simulation: MATLAB

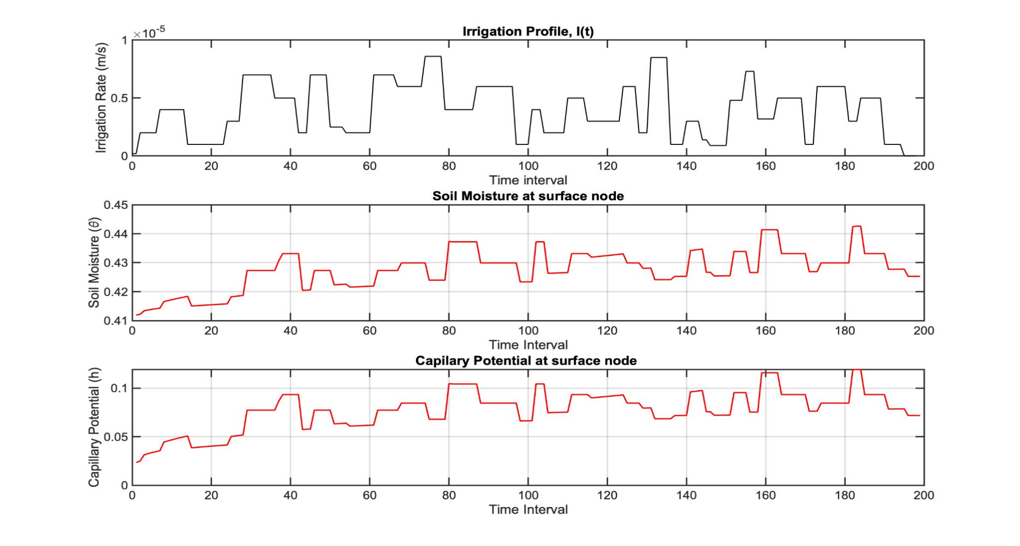
↓

Plant Data Generation: Model Training

↓ Post Midsem

Supervised & Unsupervised ML Model Training-Validation

$$\frac{dh_n}{dt} = \left(\frac{K(h_n)}{C(h_n)\cdot\Delta z}\right)\cdot\left(\frac{h_{n+1}-2h_n+h_{n-1}}{\Delta z}\right)+\left(\frac{1}{C(h_n)}\right)\cdot\frac{\partial K}{\partial h}\cdot\left(\frac{h_{n+1}-h_n}{\Delta z}\right)\cdot\left(\frac{h_{n+1}-h_n}{\Delta z}+1\right)-\frac{C_0}{C(h_n)}$$



Objective

Train & validate global LSTM model on generated data

↓

Unsupervised Learning: K-Means Clustering

↓

Train & validate local LSTM models based on operating region

↓

Ensemble learning models

Dataset:

Input feature: Irrigation Rate
Output feature: Soil Moisture

Tools & Technology used:

Python, MATLAB
Libraries/packages: Numpy, Pandas, Scikit Learn, MatPlotLib, tensorflow, optimization toolbox

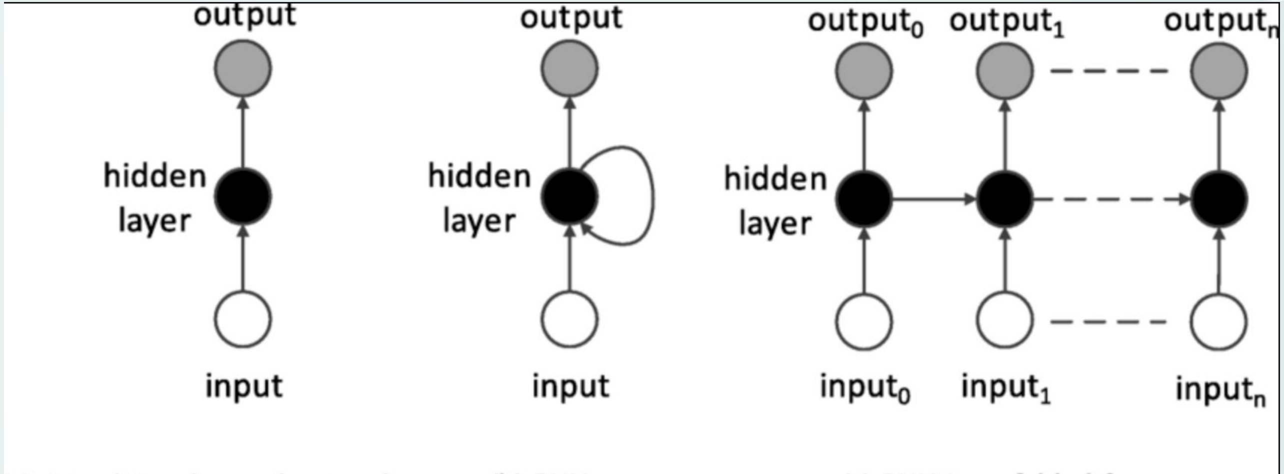
Traditional Neural Networks vs RNNs

Traditional NNs:
"Each input processed independently"

- can't remember past inputs: lacks Memory
- Image classification, tabular data prediction

RNNs:
"process one element at a time, remembering what they saw before"

- have recurrent connections: Short term memory
- Time Series Forecasting, Language Modelling



(a) A traditional neural network (b) RNN (c) RNN in unfolded form

https://www.researchgate.net/publication/330323646_Automatically_learning_usage_behavior_and_generating_event_sequences_for_black-box_testing_of_reactive_systems

Traditional NNs are **memoryless**; RNNs introduce **short-term memory**, enabling sequence learning.

Long Short Term Memory (LSTM):

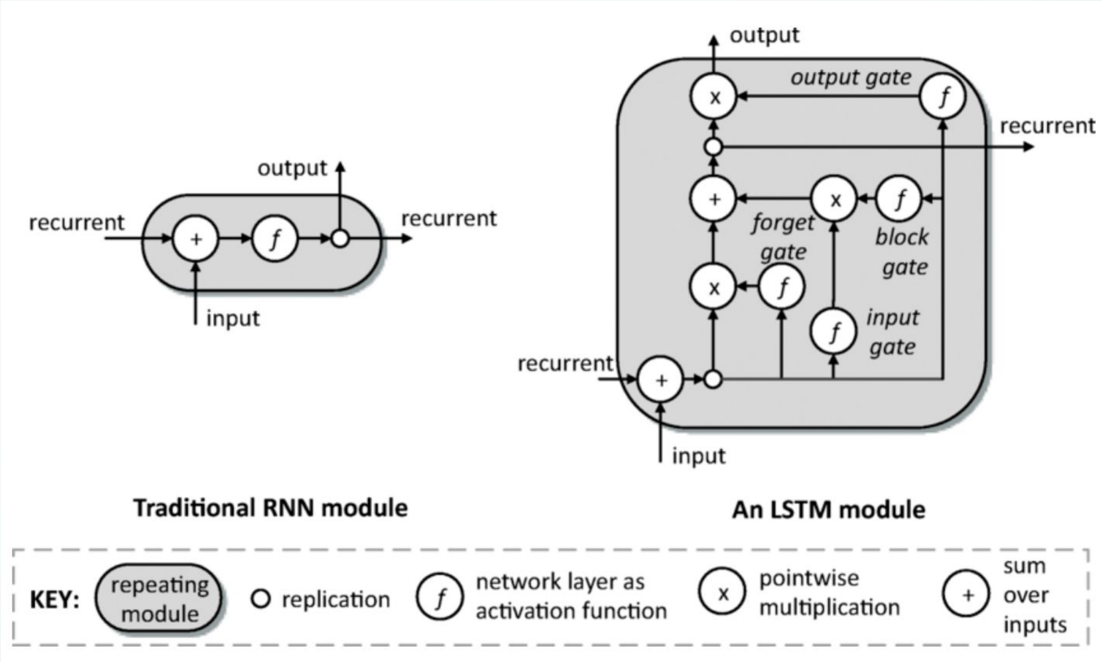
A special type of RNN designed to **remember long-term dependencies.**

.

Architecture:

- **Forget gate**: Decides what to discard
- **Input gate**: Decides what new info to store
- **Output gate**: Controls what to output
- Used in NLP, Speech recognition, Time series forecasting

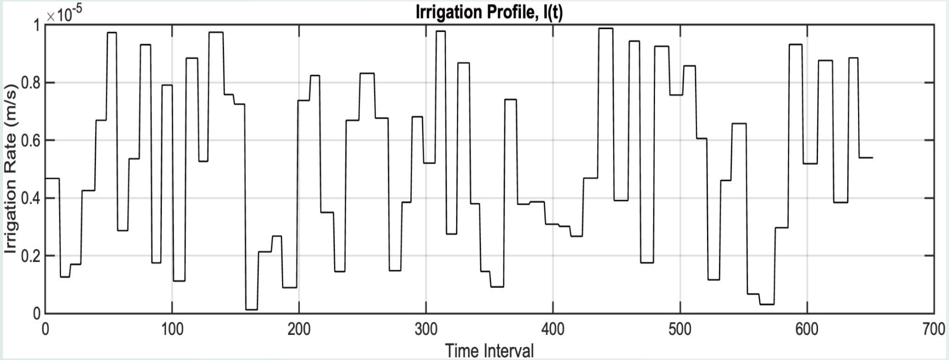
LSTMs are like "smart memory cells" that decide what to remember and forget, making them ideal for complex time-based problems.

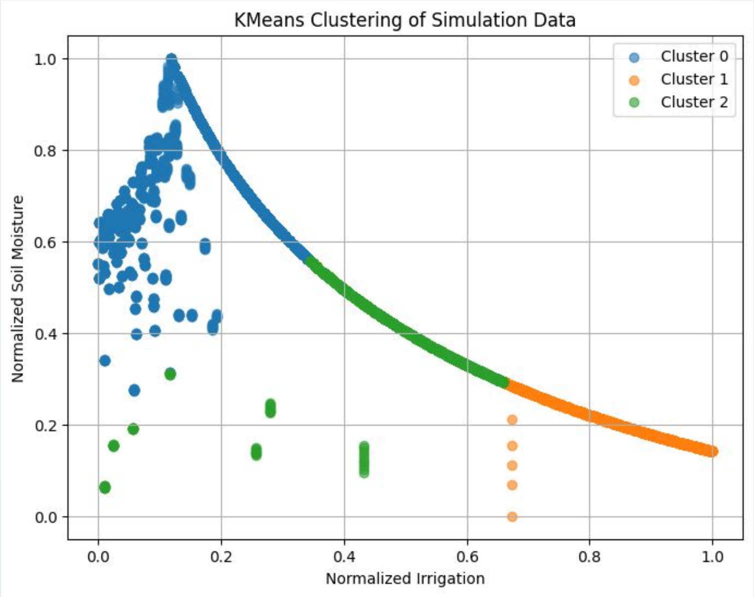


https://www.researchgate.net/figure/Traditional-RNNs-vs-LSTM-networks_fig3_330323646

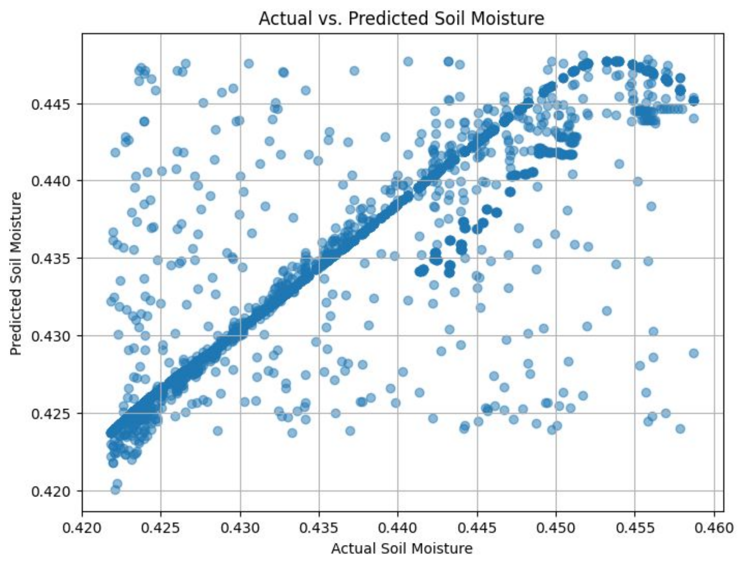
K-Means Clustering

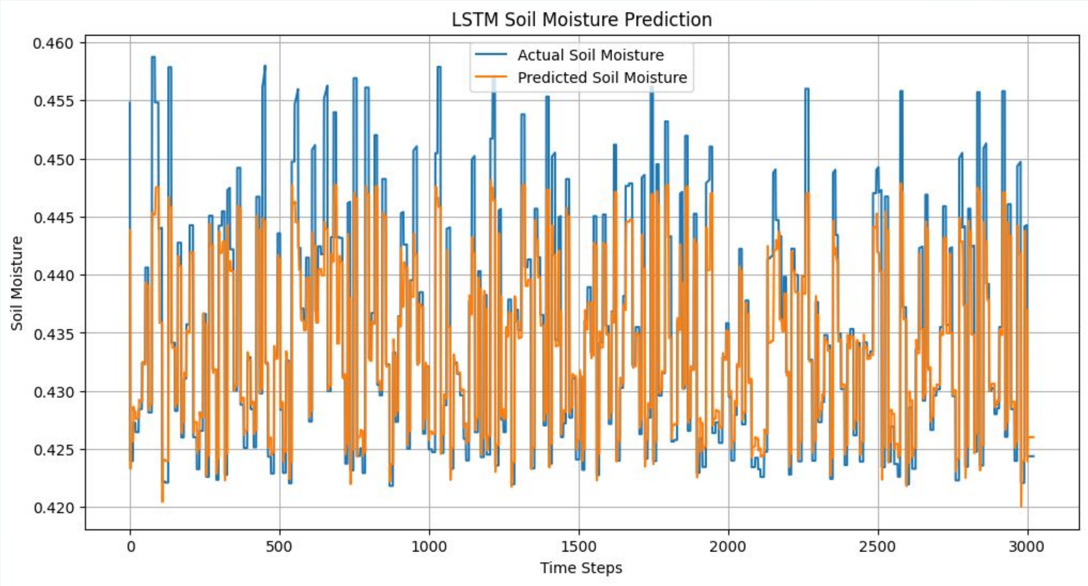
- Unsupervised learning algorithm
- Partition data into K distinct, non-overlapping clusters based on feature similarity - irrigation rate
- Captures different operating region





Results: Global LSTM





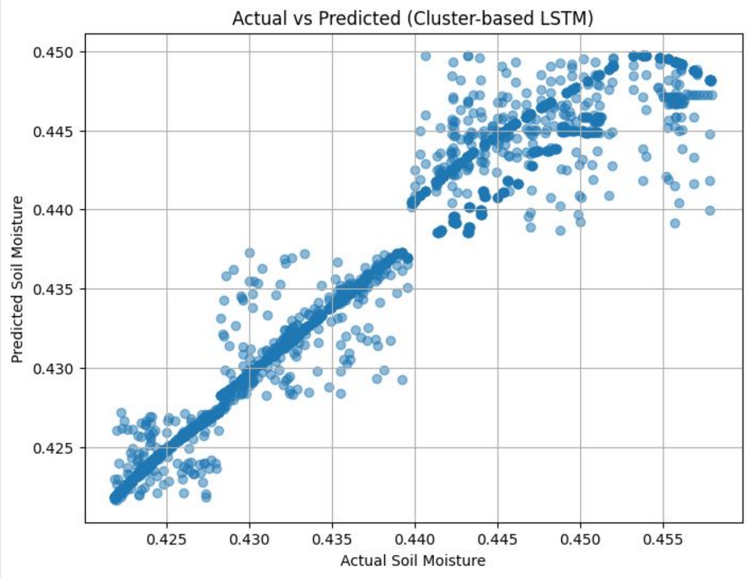
Performance Metrics:

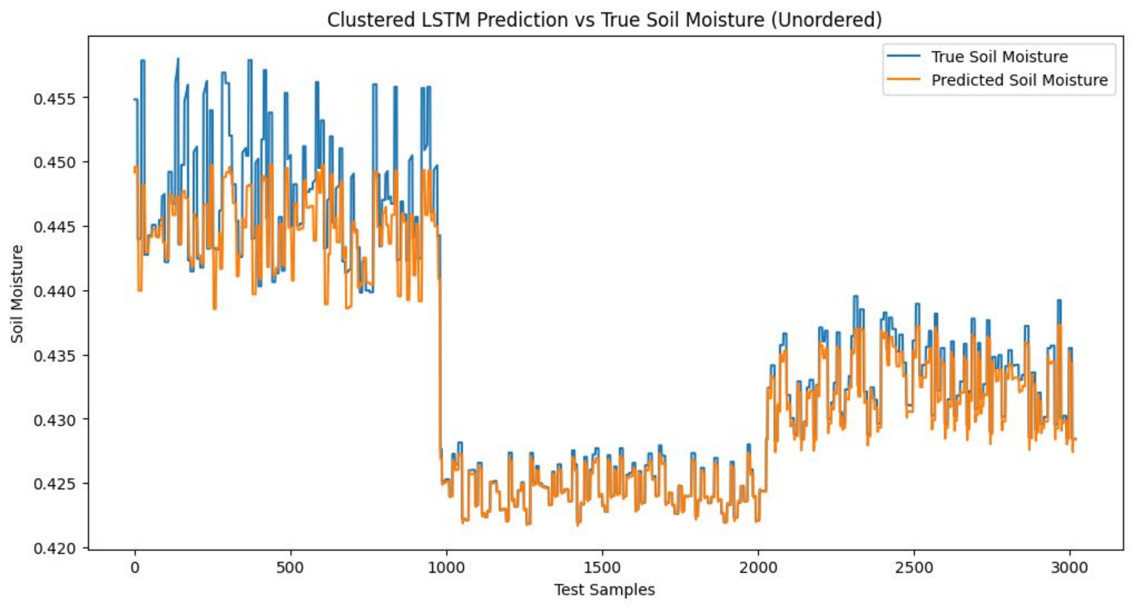
Mean Absolute Error (MAE): 0.0032

Root Mean Squared Error (RMSE): 0.0054

R^2 Score: 0.7168

Results: Clustered LSTMs

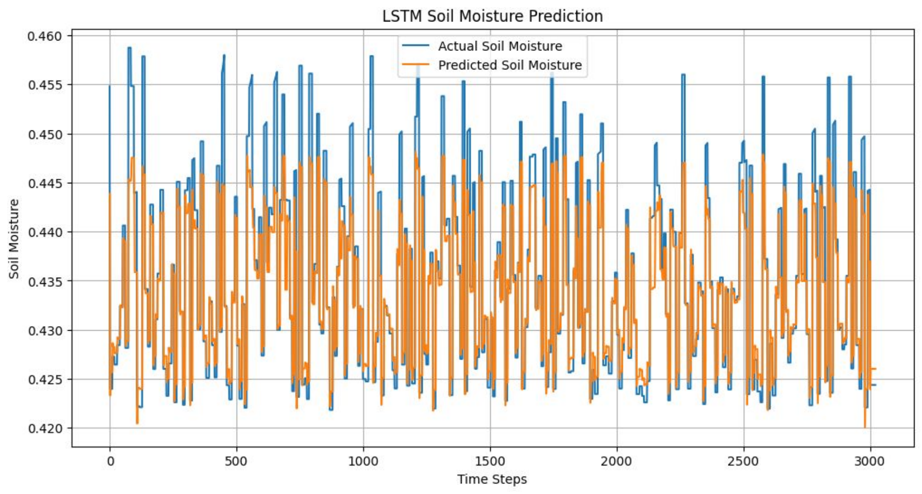




Performance Metrics:

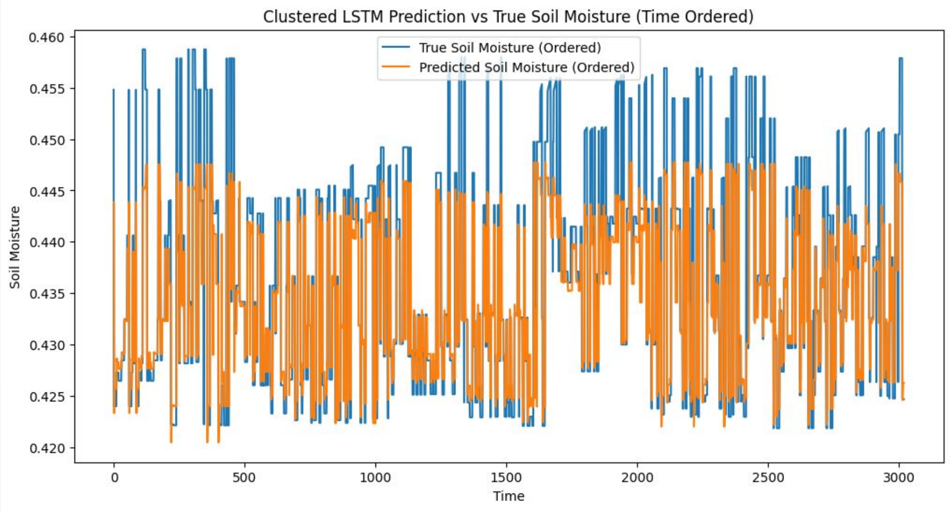
Mean Absolute Error (MAE): 0.0016 Root Mean Squared Error (RMSE): 0.0027 R^2 Score: 0.9249

Comparison: Local LSTM vs Global LSTM



Global LSTM Model

R^2 Score: 0.7168

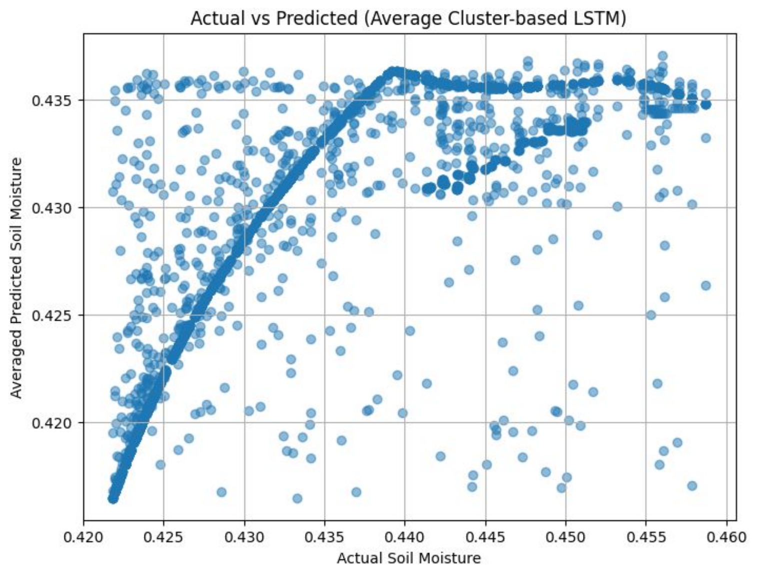


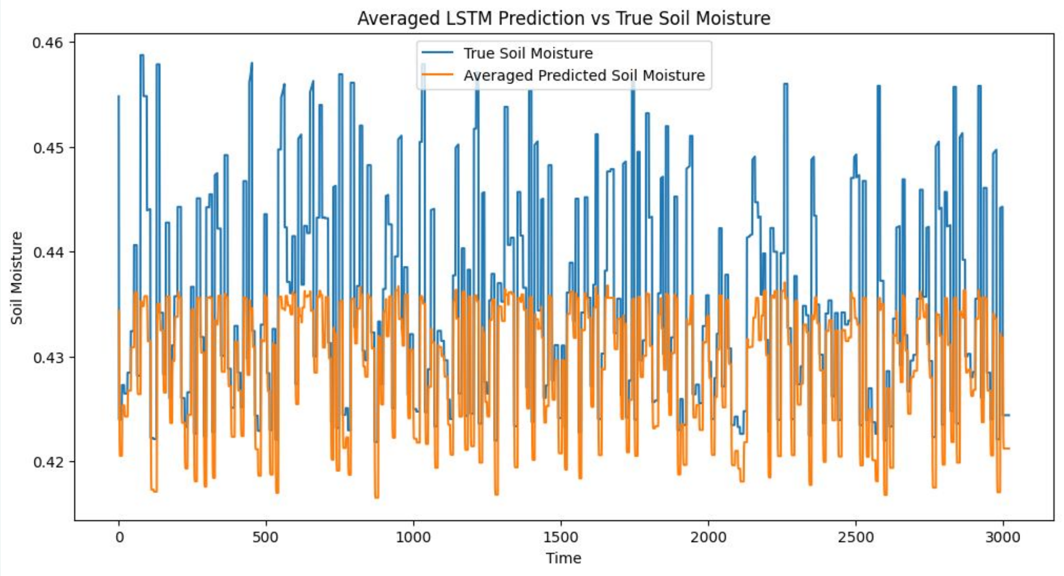
Clustered LSTM Model

R^2 Score: 0.9249

Results: Average Ensembled

$$\hat{y}_{\text{output}} = \frac{\hat{y}^{(1)} + \hat{y}^{(2)} + \hat{y}^{(3)}}{3}$$





Performance Metrics:

Mean Absolute Error (MAE): 0.0064

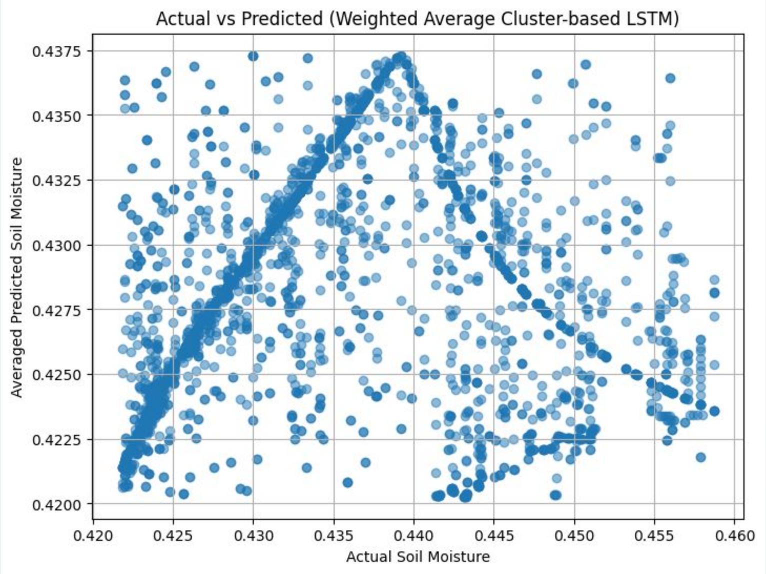
Root Mean Squared Error (RMSE): 0.0090

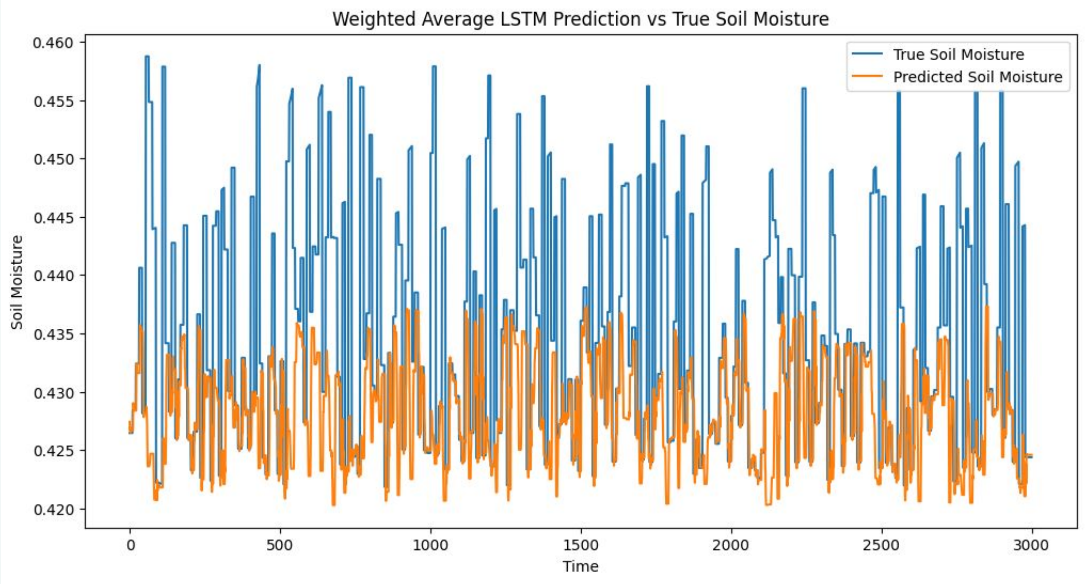
R^2 Score: 0.2024

Results: Weighted Average Ensembled

$$\hat{y}_{\text{output}} = \alpha_1 \hat{y}^{(1)} + \alpha_2 \hat{y}^{(2)} + \alpha_3 \hat{y}^{(3)}$$

$\alpha 1 + \alpha 2 + \alpha 3 = 1$





Performance Metrics:

Mean Absolute Error (MAE): 0.0081 Root Mean Squared Error (RMSE): 0.0128 R^2 Score: -0.6018

Performance Analysis

R^2 Score is statistical measure of how well the model predictions approximate the real data points.

Observation:
K-Means clustering grouped similar irrigation–soil moisture patterns into clusters, each representing a unique operating region.

The Global LSTM had to learn from the entire dataset, struggling with high variability and conflicting trends across different regions.

Models	R^2 Scores
Global LSTM Model	0.7168
Clustered LSTM Model	**0.9249**
Average Ensembled Model	0.2024
Weighted Average Ensemble Model	-0.6018

Conclusion & Future Work

In summary:

- KMeans clustering enabled segmentation of data into homogeneous regions, to handle variability in irrigation-soil moisture dynamics.
- Cluster-specific LSTM models effectively captured local temporal patterns, significantly boosting prediction accuracy (**R^2: 0.7168 → 0.9249**).
- Compared to ensemble methods, this clustering-based approach delivered the best performance, showing the importance region based modeling.

Future work:

- In the dynamic weighted ensemble model, prediction accuracy can be improved by choosing an **optimal window size**.
- A dynamic mechanism to detect changes in operating regions can help adapt window boundaries or update model weights more effectively.
- Clustering can be enhanced by including temporal features, statistical insights, or domain knowledge.

References

- Huang, Z., Liu, J., & Huang, B. (2023). Model predictive control of agro-hydrological systems based on a two-layer neural network modeling framework. International Journal of Adaptive Control and Signal Processing, 37(6), 1536-1558.
- https://www.geeksforgeeks.org/k-means-clustering-introduction/
- https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/
- https://www.researchgate.net/publication/330323646_Automatically_learning_usage_behavior_and_generating_event_sequences_for_black-box_testing_of_reactive_systems
- https://www.geeksforgeeks.org/introduction-deep-learning/

THANK YOU

For this B.Tech IIT Ropar