SRTTU

Journal of Electrical and Computer Engineering Innovations (JECEI)

Journal homepage: http://www.jecei.sru.ac.ir

JECEI

Research paper

Diverse Climatic Soil Moisture Estimation Using an Enhanced GRU Model with Multi-Source Data Augmentation

Reyhaneh Bagheri, Fatemeh Tabib Mahmoudi*, AmirHossein Gholamian

Department of Geomatics, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran.

Article Info

Article History:
Received 10 November 2025
Reviewed 03 January 2026
Revised 18 February 2026
Accepted 23 February 2026

Keywords:
Soil moisture prediction
Deep learning
Data augmentation
Gated Recurrent Unit
USCRN

*Corresponding Author's Email Address: *fmahmoudi@sru.ac.ir*

Abstract

Background and Objectives: Accurate soil moisture estimation is essential for various hydrological processes such as irrigation planning and environmental monitoring; however, prediction accuracy is often limited by sparse in-situ measurements and uncertainties in remote sensing products. This study aims to develop an enhanced soil moisture prediction framework by integrating a Gated Recurrent Unit (GRU) deep learning model with multi-source data augmentation techniques in order to evaluate its performance across diverse climatic conditions.

Methods: This study proposed an enhanced GRU deep learning framework supported by multi-source data augmentation to predict soil moisture across ten U.S. Climate Reference Network (USCRN) stations representing diverse climatic and ecological conditions. The proposed model integrates Conv1D layers, bidirectional GRUs, multi-head attention, and dense layers to capture short and long-term temporal dependencies while fusing multi-source inputs, including ISMN in-situ measurements, SMAP products, and GLDAS. Data augmentation strategies composed of noise injection, temporal warping, scaling, and window slicing were applied to expand the training dataset and reduce overfitting. Model performance was compared against a standard GRU using some of the evaluation metrics.

Results: Results demonstrate that the augmented GRU model consistently outperformed the standard GRU across all stations, with notable improvements in R^2 (up to 0.912), RMSE, and MAE. Performance gains were particularly evident in humid continental and Mediterranean climates, while regions with complex forested or semi-arid environments also benefited from data augmentation. These improvements confirm that data augmentation enhances the model's generalization under climatic variability and mitigates limitations associated with SMAP resolution and GLDAS uncertainty.

Conclusion: The integration of multi-source datasets with an augmented GRU architecture provides a reliable framework for soil moisture estimation across diverse environments. The proposed approach offers strong potential for applications in environmental monitoring, precision agriculture, and water resource management.

How to cite this paper:
R. Bagheri, F. Tabib Mahmoudi, A. M. Gholamian, "Diverse climatic soil moisture estimation using an enhanced GRU model with multi-source data augmentation," J. Electr. Comput. Eng. Innovations, 14(2): 425-434, 2026.

DOI: 10.22061/jecei.2026.12695.903

URL: https://jecei.sru.ac.ir/article_2540.html

Introduction

Soil moisture is one of the key variables in modeling hydrological processes, managing water resources, and improving irrigation efficiency [1]-[4]. Accurate estimation of soil moisture is particularly important at large agricultural scales. One of the main challenges in water resource management and agriculture is the precise estimation and prediction of root-zone soil moisture. Soil moisture not only plays a crucial role in plant growth and determining water requirements but also is highly important in modeling the hydrological cycle, drought prediction, and environmental crisis management [5]-[8].

Most traditional methods of estimating the soil moisture rely on ground stations or field measurements [9]-[11]. These methods can only give information about soil moisture near the surface in small, specific areas. Because of the changing conditions over time and space, traditional models often struggle to provide accurate soil moisture estimations. Traditional methods for measuring soil moisture on a large scale are often difficult, costly, and limited to local stations; therefore, faster and more accurate approaches such as machine learning and deep learning models using remote sensing data have received considerable attention in recent years [12], [13].

Soil moisture estimation methods based on remote sensing data can be categorized into two main types based on the utilized data: optical and microwave [14]-[16]. In using optical remote sensing, Gojiya et al. demonstrated the application of Artificial Neural Networks (ANNs) with Landsat and Sentinel-2 derived spectral indices to estimate surface soil moisture in the semi-arid region of India [15]. Singh and Gaurav, developed a fully connected feed-forward Artificial Neural Network to estimate high-resolution surface soil moisture by fusing multi-sensor satellite data from Sentinel-1, Sentinel-2 and SRTM digital elevation model [17] Lamichhane et al. summarized the progress in using machine learning algorithms with remote sensing data for soil moisture estimation. Their analysis reveals that Random Forest, Support Vector Regression, and Artificial Neural Networks are the most frequently used and best-performing algorithms in soil moisture estimation. The integration of multi-source remote sensing data, such as combining optical, thermal, and microwave sensors, consistently enhances model accuracy compared to single-source approaches [18]. In recent years, there has been more research on using high-resolution remote sensing data, both optical and microwave, especially for managing environmental issues [10], [19]-[21].

Deep learning techniques have shown great potential in identifying complex nonlinear relationships within environmental data [5]. Among deep learning methods, Recurrent Neural Networks (RNNs), particularly the Gated Recurrent Unit (GRU), demonstrate high performance in time series prediction due to their ability to preserve temporal dependencies and mitigate the vanishing gradient problem [22]. Previous studies have shown that GRU can improve the accuracy of soil moisture prediction, and when combined with other architectures such as Transformer or long short-term memory (LSTM), the model's ability to capture both short-term and long-term relationships is further enhanced [5]. For instance, Baumberger et al. used the GRU model to predict the time series of soil temperature and moisture and found that this model had higher accuracy than LSTM and ANN, while also simulating the physical relationships between environmental features and soil conditions [23]. Zheng et al. using a hybrid GRU–Transformer model, demonstrated that combining the GRU's capability to extract short-term dependencies with the Transformer's ability to identify long-term relationships enhances the prediction performance of soil moisture in the root zone [24]. Wang et al. improved the GRU by applying a parameter optimization algorithm, which increased prediction accuracy over short time intervals [22]. Moreover, Munaganuri et al. showed that GRU combined with LSTM performs well in agricultural yield prediction using environmental data and can enhance agricultural productivity and sustainability [25]. Ahmed et al. developed a sophisticated deep learning hybrid model for daily surface soil moisture prediction in the Australian Murray Darling Basin. Their approach uniquely integrated MODIS satellite data, ground-based meteorological observations, and synoptic-scale climate indices. The model combined Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for data decomposition, Convolutional Neural Networks for feature extraction, and a GRU network for forecasting [26].

In addition to GRU, other deep learning approaches have also been utilized for soil moisture estimation. Vyas et al. identified spatial–temporal relationships among different points using a dynamic graph neural network [27]. Wang et al. proposed the SMF4WSN model for soil moisture prediction in wireless sensor networks, which improved prediction accuracy [28]. Grubišić et al. demonstrated that models enhanced with the attention mechanism have a higher capability to simulate the complex dynamics of soil moisture [29]. The study conducted by Heggy et al. introduced an integrated deep learning framework designed to predict soil moisture and fill temporal gaps in observational datasets by leveraging LSTM networks and convolutional components. Their model successfully reconstructed missing soil moisture values and demonstrated superior performance compared to traditional interpolation and statistical techniques by effectively learning nonlinear hydrological patterns from remote sensing and in-situ datasets [30]. In addition to the above studies, several foundational and state-of-the-art works provide essential context for soil moisture estimation. Recent advances emphasize hybrid deep learning models [31], and cross-climatic validation using ISMN and global datasets [32].

Moreover, emerging research highlights the importance of data augmentation in time-series deep learning [33], [34], yet its application to soil moisture prediction remains largely unexplored. These gaps further justify the need for a multi-source, augmentation-enhanced GRU framework evaluated across diverse climates, as proposed in the present study.

Table 1 summarizes key studies in soil moisture estimation to facilitate the comparison of the previous studies. However, the literature still suffers from several key limitations: (i) limited integration of multi-source datasets such as SMAP, GLDAS, and in-situ observations; (ii) minimal use of data augmentation techniques to overcome sparse ground measurements and reduce model overfitting, and (iii) insufficient cross-climatic evaluation, with most studies focusing on a single station or climatic region.

The main objective of this study is to investigate and compare the performance and accuracy of the GRU network in estimating the soil moisture under standard and improved conditions using data augmentation. As a recurrent neural network, GRU has the capability to preserve temporal dependencies and reduce the vanishing gradient problem in time series data.

In this study, efforts have been made to utilize this network appropriately and to enhance prediction accuracy over a large study area, ensuring that the obtained results have acceptable accuracy for applications in water resources management and sustainable agriculture. The use of data augmentation techniques, by increasing the diversity and volume of training data, not only reduces overfitting but also improves the model's generalization ability and enhances prediction accuracy under real-world conditions. The main contributions of this study are listed below:

- The integration of a multi-input GRU architecture combining Conv1D, bidirectional GRU, and attention mechanisms to improve temporal feature extraction.
- The systematic use of multi-source data augmentation (noise injection, temporal warping, scaling, and window slicing) applied to soil moisture prediction.
- The proposed framework is evaluated across diverse climatic and ecological conditions within the United States to assess robustness and generalizability.

Table 1: Literature analysis and comparison

Study	Datasets Used	Model / Algorithm	Climate / Study Area	Limitations / Research Gaps
Gojiya et al. (2024)	Landsat, Sentinel-2	ANN	Semi-arid (India)	Single climate, no temporal modeling, no multi-source fusion
Singh & Gaurav (2023)	Sentinel-1, Sentinel-2, SRTM	Fully connected ANN	Agricultural areas	Lacks temporal dependency modeling; no augmentation
Lamichhane et al. (2025)	Multi-source RS	RF, SVR, ANN	Multiple regions	No deep temporal models; doesn't address data scarcity
Baumberger et al. (2025)	In-situ, meteo data	GRU	Europe	No multi-source RS; single region; no augmentation
Wang et al. (2023)	Remote sensing	Improved WCM + DL	Agricultural	Limited benchmarking; no multi-source inputs
Zheng et al. (2024)	In-situ + RS	GRU–Transformer hybrid	Agricultural	No data augmentation; limited climate variability
Ahmed et al. (2021)	MODIS, SILO, climate indices	CNN + GRU + CEEMDAN	Australia	No cross-climatic testing; no augmentation; single-country dataset
Orth (2021)	In-situ + ML	ML regression models	Global	Low-res, lacks deep learning temporal modeling
Wang et al. (2024)	Soil station data	Improved GRU	Agricultural	No multi-source RS; no augmentation; limited generalization
Grubišić et al. (2024)	NDVI + RS	Attention-based DL	Multiple	No multi-input GRU; limited temporal fusion
ElSaadani et al. (2021)	RS + in-situ	LSTM–CNN	Multi-region	Limited climate diversity; no augmentation methods
Present Study (2025)	SMAP, GLDAS, ISMN (multi-source)	Conv1D + BiGRU + Attention + Data Augmentation	Ten USCRN stations representing diverse climates	Addresses gaps in generalizability, multi-source fusion, deep temporal modeling, and data scarcity

The remainder of this paper is organized as follows. Section 2 describes the study area and the utilized multi-source datasets. Section 3 presents the methodology, detailing the data augmentation strategies and the architecture of the proposed GRU-based model. Section 4 reports the experimental setup, evaluation metrics, and comparative results between the augmented GRU and the standard GRU models, followed by an in-depth discussion of model performance across different climatic conditions. Section 5 concludes the study by summarizing the key findings and highlighting potential directions for future research.

Dataset and Study Area

In this study, three complementary datasets were employed to estimate soil moisture:

- **ISMN (International Soil Moisture Network):** Ground-based in-situ observations used as reference for model evaluation.
- **SMAP (Soil Moisture Active Passive):** Satellite-based Level 3 soil moisture products with global coverage and ~36 km spatial resolution.
- **GLDAS (Global Land Data Assimilation System):** Reanalysis data providing meteorological and hydrological variables such as precipitation and temperature.

These datasets were temporally aligned and resampled to a common daily time step to ensure consistency across multiple sources. The study period spans the entire years of 2024 and 2025.

All datasets underwent standard preprocessing steps composed of merging, normalization (for preventing data leakage) and 10-days sequences creation. This study analyzes ten U.S. Climate Reference Network (USCRN) stations, chosen to represent a wide range of climatic and ecological conditions across the United States (Fig. 1). As summarized in Table 2, the stations encompass humid continental agricultural regions (e.g., Aberdeen, Lincoln, Oakley), semi-arid grasslands and high-elevation environments (e.g., John Day, Muleshoe, Los Alamos), dense forest and mountainous areas with high precipitation (e.g., Darrington), and Mediterranean coastal and Central Valley ecosystems (e.g., Bodega, Merced). This diversity provides a comprehensive basis for evaluating the robustness and generalizability of the proposed GRU-based soil moisture estimation framework under varying climatic regimes and environmental complexities [35]. These stations, situated in minimally disturbed environments, provide high-quality in-situ data, integrated with SMAP, GLDAS, and ISMN datasets to monitor soil moisture dynamics by using remote sensing data.

The diversity in climate, soil, and land cover characteristics allows the study to assess the performance of deep learning models, especially the GRU network enhanced with data augmentation, for accurate and reliable soil moisture estimation. The integration of multi-source datasets (ISMN, GLDAS, SMAP) further strengthens model robustness and generalizability across varying environmental conditions.

Table 2: Summary of USCRN stations: Coordinates, site type, climate, and study purpose

Station Name	Latitude / Longitude	Site Type	Climate Classification	Study Purpose / Relevance
Avondale-2-N, PA	40.0°N, 75.9°W	Agricultural–Forested Mixed Landscape	Humid Continental	Evaluating soil moisture dynamics in mixed agricultural–riparian settings; SMAP and ISMN validation.
Darrington-21-NNE, WA	48.5°N, 121.6°W	Mountainous Forest	Oceanic / Wet Maritime	Assessing model performance in dense forest canopies and high-precipitation regimes using GLDAS inputs.
John-Day-35-WNW, OR	44.7°N, 119.6°W	Semi-Arid Grassland	Semi-Arid Steppe	Testing robustness under dry climates and sparse vegetation; drought impact analysis.
Aberdeen-35-WNW, SD	45.7°N, 98.5°W	Agricultural Plains	Humid Continental	Soil moisture prediction in crop-dominated landscapes with strong seasonal variability.
Lincoln-11-SW, NE	40.7°N, 96.9°W	Cropland (Corn–Forage)	Humid Continental	Evaluating GLDAS-driven soil moisture variability and agricultural hydrological behavior.
Muleshoe-19-S, TX	34.0°N, 102.7°W	Dry Grassland / Wildlife Refuge	Semi-Arid	Validating model performance in arid regions with rapid moisture fluctuations.
Los-Alamos-13-W, NM	35.8°N, 106.5°W	High-Elevation Forest (Ponderosa Pine)	Semi-Arid / Highland	Examining impacts of elevation and rocky soils on moisture predictability and ISMN integration.
Bodega-6-WSW, CA	38.3°N, 123.1°W	Coastal Mediterranean	Mediterranean	Evaluating soil moisture behavior influenced by fog events and coastal vegetation.
Oakley-19-SSW, KS	39.0°N, 100.9°W	Central Plains Agriculture	Humid Continental	Investigating model sensitivity to crop-driven soil moisture variations.
Merced-23-WSW, CA	37.1°N, 120.7°W	Central Valley Wetlands/Agriculture	Mediterranean	Assessing soil moisture dynamics in wetland–agricultural mosaics; testing model generalization.

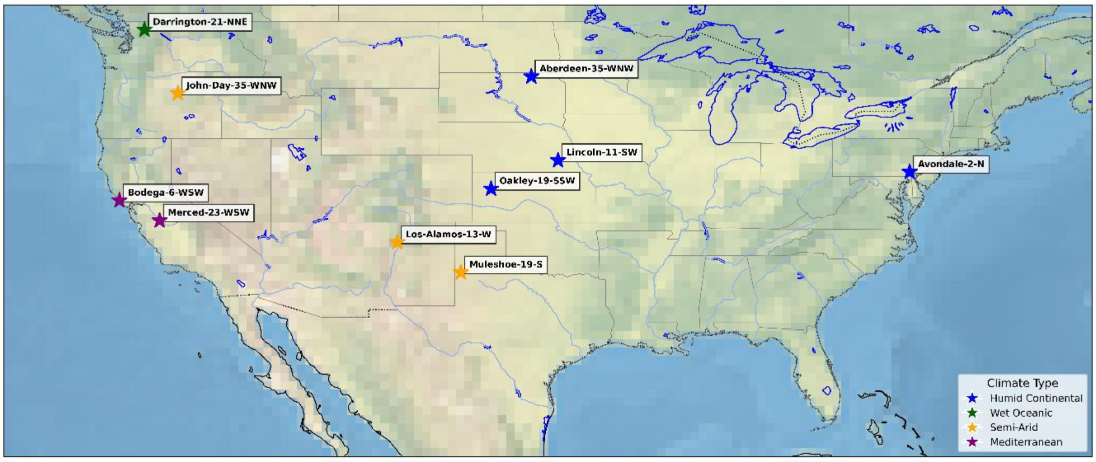

Fig. 1: Distribution map of the utilized USCRN stations by climate types.

Method

Figure 2 illustrates the key steps of the proposed methodology in using multi-input data sets in GRU model together with data augmentation for soil moisture estimation.

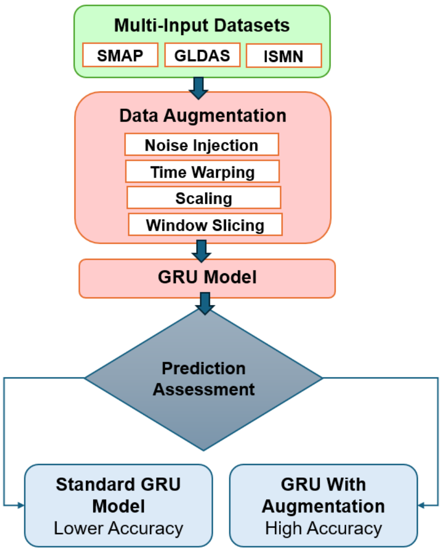

Fig. 2: Structure of the proposed soil estimation method.

A. Data Augmentation

To enhance the robustness of the model and mitigate overfitting caused by the limited availability of soil moisture data, several data augmentation techniques were employed:

- **Noise Injection:** Small Gaussian noise was added to the input sequences to improve generalization and reduce sensor uncertainty.
- **Time Warping:** Random distortions in the temporal dimension were introduced to simulate natural variability in soil moisture dynamics.
- **Window Slicing:** Fixed-length subsequences were extracted from the longer time series to increase the number of training samples.
- **Scaling:** Random scale factors were applied to input features to simulate differences in soil and vegetation responses.

These augmentation methods increased the effective size of the training data, allowed the model to generalize better, and improved resilience against seasonal and meteorological fluctuations.

B. Model Architecture

A GRU network was implemented due to its efficiency in capturing temporal dependencies in time-series data while requiring fewer parameters than an LSTM. As illustrated in Fig. 3, the architecture consists of the following parts:

- **Input Layer:** Multi-source input data consists of SMAP, GLDAS, and ISMN.
- **GRU Layers:** One or more stacked GRU layers to capture long-term and short-term dependencies in soil moisture variability.
- **Dense Layer:** Fully connected layer to map hidden states into soil moisture prediction.
- **Activation Function:** The ReLU function was applied to enhance non-linearity and prevent vanishing gradients.

The proposed model is a multi-input GRU-based neural network designed to fuse temporal patterns from trainable parameters, balancing complexities and computational efficiency. Three parallel input layers process the normalized 10-day time-series data from each source independently. A 1D convolutional layer (Conv1D) with 64 filters, a kernel size of 3, and ReLU activation function extracts local temporal patterns, such as short-term fluctuations in soil moisture. A Bidirectional GRU layer with 128 units capturing both forward and backward temporal dependencies to model long and short-term dynamics of ISMN, SMAP, and GLDAS data, leveraging convolutional, recurrent, and attention mechanisms for accurate soil moisture prediction. Layer normalization stabilizes training by normalizing feature scales, while dropout (rate=0.3) prevents overfitting by randomly deactivating units, crucial for the limited dataset size.

The processed sequences from all sources are concatenated, resulting in a combined sequence integrating multi-source information. ReLU activation throughout the architecture ensures non-linearity, mitigating vanishing gradient issues, while the linear output aligns with the continuous nature of soil moisture values.

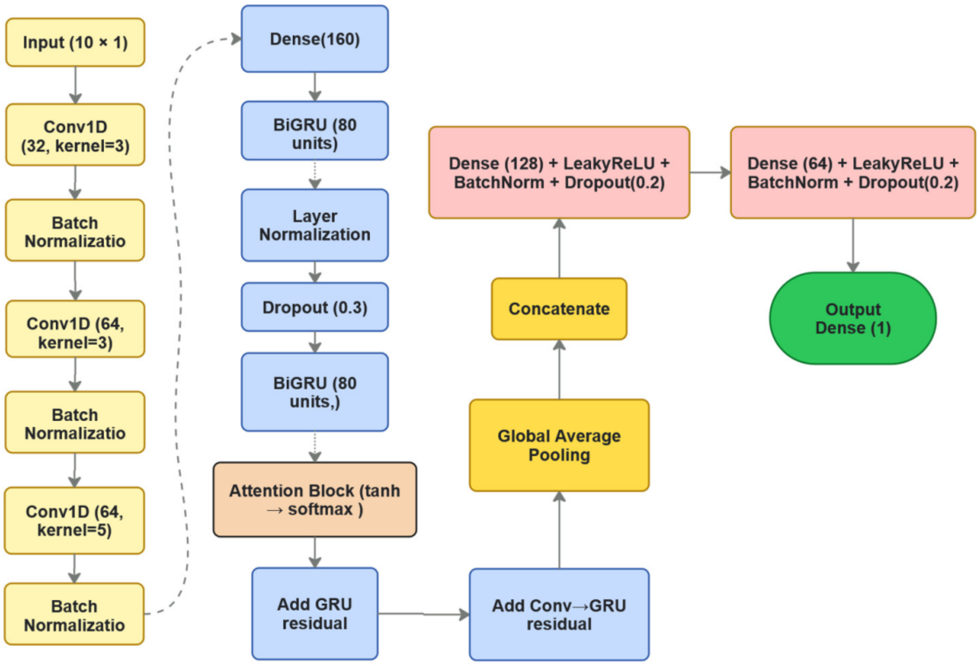

Fig. 3: Proposed enhanced GRU model architecture.

Results and Discussion

A. Implementation Details

The model was implemented and trained on an NVIDIA Tesla T4 GPU using the Adam optimizer with an initial learning rate of 1×10^{-4}. Training was performed for 50 epochs with a batch size of 4, using MSE as the loss function and MAE as the performance metric. To enhance reliability and ensure stable convergence, three strategies were applied: *ReduceLROnPlateau* dynamically decreased the learning rate (factor = 0.2, patience = 7, minimum LR = 1×10^{-6}) when no improvement in validation MAE was observed; *ModelCheckpoint* preserved the best-performing model based on validation MAE; and *EarlyStopping* halted training upon stagnation of performance, restoring the optimal weights. These procedures improved long-term training stability, mitigated overfitting, and ensured statistically robust results. The training loss decreased from 1.53 at initialization to 0.026 in the final epochs, confirming consistent convergence and the effectiveness of the adopted optimization and stabilization mechanisms.

To prevent overfitting, early stopping was applied by monitoring the validation loss, stopping training if it did not improve for 10 consecutive epochs, and restoring the best weights. In addition, model checkpointing was used to save the best-performing model during training based on the lowest validation loss, ensuring that the optimal model was preserved.

B. Evaluation Metrics

Model performance was assessed on the validation set using the following metrics to evaluate both errors and correlations:

- **Pearson Correlation Coefficient (R):** Assesses the linear correlation between predicted and observed values.

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

- **Root Mean Squared Error (RMSE):** Represents the square root of MSE for better interpretability.

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

- **Mean Absolute Error (MAE):** defined as the average sum of the absolute differences between the actual value and the predicted value.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y_i}\right| \quad (3)$$

- **Coefficient of Determination (R^2):** with a threshold of ≥0.75 for station acceptance.

$$R^2 = \frac{1-\sum(y_i-\hat{y_i})^2}{\sum(y_i-\bar{y})^2} \quad (4)$$

C. Results and Comparison

To investigate the impact of data augmentation on the performance of the GRU model for soil moisture estimation, the results of ten stations (with and without data augmentation) are presented in Tables 3 and 4.

Table 3: Assessment of GRU model with data augmentation

Stations	Loss	MAE	R	R²	RMSE
Merced-23-WSW	0.002	0.029	0.991	0.9746	0.046
Bodega-6-WSW	0.004	0.034	0.983	0.9601	0.066
Darrington-21-NNE	0.004	0.044	0.987	0.9637	0.065
Los-Alamos-13-W	0.003	0.033	0.984	0.9658	0.054
Aberdeen-35-WNW	0.003	0.034	0.984	0.9621	0.052
Avondale-2-N	0.011	0.048	0.949	0.8889	0.103
Oakley-19-SSW	0.007	0.038	0.950	0.8981	0.083
Muleshoe-19-S	0.007	0.024	0.925	0.8501	0.087
Lincoln-11-SW	0.018	0.075	0.882	0.7535	0.136
John-Day-35-WNW	0.015	0.082	0.892	0.8225	0.157

The results obtained from the 10 stations show that the GRU model with data augmentation exhibits a substantial improvement over the standard GRU model across all error and correlation metrics. The average indicators reveal that the Loss value decreased from 0.0257 to 0.0075, which corresponds to a 70.8% improvement. The MAE and RMSE metrics also decreased by 66.4% and 47.9%, respectively; this reduction indicates higher model accuracy and a significant decrease in the gap between the observed and predicted values. Moreover, the correlation indices also improved; the R^2 increased from 0.6838 to 0.9039, representing a 32.2% improvement and indicating that the model explains more than 90% of the variance in the actual data. The 5.9% increase in the Pearson coefficient further reflects the strengthening of the linear relationship between the model output and the measured data.

Table 4: Assessment of standard GRU without data augmentation

Stations	Loss	MAE	R	R²	RMSE
Merced-23-WSW	0.0159	0.0891	0.9267	0.7968	0.1324
Bodega-6-WSW	0.0245	0.1217	0.9421	0.7531	0.1521
Darrington-21-NNE	0.0298	0.1517	0.9135	0.5214	0.1816
Los-Alamos-13-W	0.0229	0.1367	0.8012	0.6401	0.1682
Aberdeen-35-WNW	0.0241	0.1279	0.8698	0.6376	0.1672
Avondale-2-N	0.0531	0.1645	0.8745	0.5136	0.2612
Oakley-19-SSW	0.0274	0.1218	0.8739	0.6841	0.1391
Muleshoe-19-S	0.0209	0.1901	0.9378	0.8653	0.1578
Lincoln-11-SW	0.0239	0.1189	0.9351	0.7528	0.1572
John-Day-35-WNW	0.0148	0.0872	0.9196	0.6728	0.1116

These findings confirm that employing data augmentation enhances the model's ability to learn complex temporal patterns in soil moisture and increases its robustness against data scarcity and temporal non-uniformity. At the station level, interesting patterns are also observed. The greatest improvement is recorded in humid stations such as Darrington-21-NNE, where RMSE decreased by more than 100%. This is likely due to the high variability in moisture conditions in these regions and the positive effect of augmentation in simulating rainfall events and soil saturation. In contrast, stations in arid regions such as Muleshoe-19-S and Avondale exhibit smaller improvements, and in some metrics even slight declines. This can be attributed to the limited diversity in the raw data, the uniformity of soil moisture conditions, and the model's lower sensitivity to augmentation in low-precipitation areas.

Therefore, from a hydrological perspective, the results indicate that the GRU with data augmentation model performs very well in humid climates. In these regions, the mean R^2 is 0.9586, and the mean RMSE is 0.0651, demonstrating the model's strong capability to reproduce pronounced soil-moisture fluctuations, heavy rainfall events, and rapid water infiltration. Under such conditions, data augmentation has helped the model better represent extreme hydrological events such as floods or prolonged saturation periods.

In semi-arid climates, the model’s performance remains good but more variable. A mean R^2 of 0.9041 and the RMSE of 0.0917 show that the model has been able to reconstruct the seasonal patterns of soil moisture with reasonable accuracy. However, in stations such as John-Day, which experience prolonged drought periods, a higher RMSE (0.157) is observed. In such situations, augmentation alone is insufficient to simulate moisture events, and it is recommended to incorporate auxiliary data such as precipitation, vegetation cover, or soil temperature.

In arid climates, the model performs weaker than in other groups. A mean R^2 of 0.8889 and a mean RMSE close to 0.1032 indicate that due to the low variability in soil moisture and the occurrence of sudden rainfall events, the available data are not sufficient to capture the actual system behavior. In stations such as Avondale, augmentation has had a moderate effect, and the model is more suitable for applications such as drought monitoring or forecasting long-term trends rather than for precise modeling of short-term fluctuations.

Conclusion

This study developed an enhanced GRU-based deep learning framework for soil moisture estimation by integrating multi-source datasets, including SMAP, GLDAS, and ISMN, with a data augmentation strategy. The results obtained from ten USCRN stations demonstrated that the augmented GRU model consistently outperformed the standard GRU across all climatic regions examined. Higher values of R^2, along with reductions in RMSE, MAE, and loss, confirm that combining GRUs model and data augmentation significantly improves the model’s robustness and predictive capability. In addition, the results showed that the model performed best in humid continental and Mediterranean climates, and even under more complex conditions such as humid forests or semi-arid climates, it provided higher accuracy than GRU. These findings highlight the importance of incorporating data augmentation when dealing with limited or unevenly distributed soil moisture observations, as it enhances generalization and mitigates the limitations imposed by coarse satellite resolution and environmental variability.

The contributions of this work lie in introducing a multi-input hybrid GRU architecture capable of capturing both short- and long-term temporal dependencies, applying a comprehensive augmentation pipeline to enrich training sequences, and evaluating model performance across diverse climatic and ecological conditions.

While the proposed framework offers substantial improvements in soil moisture prediction, several limitations should be acknowledged. The performance of the model remains partly constrained by the coarse spatial resolution of SMAP products and the inherent uncertainties present in GLDAS inputs, climatic bias, and specific environmental conditions such as local soil properties and vegetation cover which may introduce biases in regions with heterogeneous land cover or complex terrain. The model also relies heavily on the quality and representativeness of the available in-situ measurements, and its sensitivity to augmentation parameters may affect performance in climates characterized by abrupt moisture fluctuations. Additionally, the current architecture focuses on temporal modeling without explicitly incorporating spatial dependencies, which may limit its applicability for regional or continental-scale soil moisture mapping.

These limitations highlight the need for future research to integrate higher-resolution remote sensing products, extending the model to more diverse or extreme climatic regions, and exploring hybrid architectures that integrate transformers, graph neural networks, or physics-informed constraints. Additionally, expanding the range of augmentation techniques and assessing the transferability of the model across regions and continents may further strengthen its applicability for large-scale soil moisture monitoring.

Funding

This work was supported by Shahid Rajaee Teacher Training University under grant number 1404 / 394107.

Author Contributions

Reyhaneh Bagheri has the following rolls in this manuscript: Data capturing, Formal analysis; Methodology implementation. Fatemeh Tabib Mahmoudi as the corresponding author has the following responsibilities: Supervision; Results’ validation; Visualization; Writing - original draft; Writing – review. AmirHossein Gholamian as the third author was responsible for system initialization and data analysis.

Acknowledgment

Authors acknowledge the dept. of Geomatics Engineering of Shahid Rajaee Teacher Training University for their supports.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
USCRN	U.S. Climate Reference Network
ISMN	International Soil Moisture Network
SMAP	Soil Moisture Active Passive
GLDAS	Global Land Data Assimilation System
MAE	Mean Absolute Error

References

[1] Q. Han, Y. Zeng, L. Zhang, C. Wang, E. Prikaziuk, Z. Niu, "Global long term daily 1 km surface soil moisture dataset with physics informed machine learning," Sci. Data, 10(1): 101, 2023.

[2] N. Madhukumar, E. Wang, Y. Everingham, W. Xiang, "Hybrid transformer network for soil moisture estimation in precision irrigation," IEEE Access, 12: 48898-48909, 2024.

[3] G. Patrizi, A. Bartolini, L. Ciani, V. Gallo, P. Sommella, M. Carratù, "A virtual soil moisture sensor for smart farming using deep learning," IEEE Trans. Instrum. Meas., 71: 1-11, 2022.

[4] G. Nijaguna, D. Manjunath, M. Abouhawwash, S. S. Askar, D. K. Basha, J. Sengupta, "Deep learning-based improved WCM technique for soil moisture retrieval with satellite images," Remote Sens., 15(8): 2005, 2023.

[5] Y. Wang, L. Shi, Y. Hu, X. Hu, W. Song, L. Wang, "A comprehensive study of deep learning for soil moisture prediction," Hydrol. Earth Syst. Scie. Discuss., 28: 917-943, 2023.

[6] J. Y. Seo, S. I. Lee, "Probabilistic evaluation of drought propagation using satellite data and deep learning model: from precipitation to soil moisture and groundwater," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 16: 6048-6061, 2023.

[7] R. Orth, "Global soil moisture data derived through machine learning trained with in-situ measurements," Sci. Data, 8(1): 1-14, 2021.

[8] L. Li, Y. Dai, W. Shangguan, Z. Wei, N. Wei, Q. Li, "Causality-structured deep learning for soil moisture predictions," J. Hydrometeorol., 23(8): 1315-1331, 2022.

[9] T. Schmidt, M. Schrön, Z. Li, T. Francke, S. Zacharias, A. Hildebrandt, "Comprehensive quality assessment of satellite-and model-based soil moisture products against the COSMOS network in Germany," Remote Sens. Environ., 301: 113930, 2024.

[10] J. Peng, A. Loew, "Recent advances in soil moisture estimation from remote sensing," Water, 9(7): 530, 2017.

[11] S. Kanzari, S. B. Mariem, S. Ghannem, S. Mouelhi, H. Ghazouani, B. B. Nouna, "A framework based on mechanistic modelling and machine learning for soil moisture estimation," Discover Soil, 2(1): 56, 2025.

[12] A. Fluhrer, T. Jagdhuber, C. Montzka, M. Schumacher, H. Alemohammad, A. Tabatabaeenejad, "Soil moisture profile estimation by combining P-band SAR polarimetry with hydrological and multi-layer scattering models," Remote Sens. Environ., 305: 114067, 2024.

[13] Q. Liu, Z. Wu, N. Cui, X. Jin, S. Zhu, S. Jiang, "Estimation of soil moisture using multi-source remote sensing and machine learning algorithms in farming land of Northern China," Remote Sens., 15(17): 4214, 2023.

[14] D. Lakra, S. Pipil, P. K. Srivastava, S. K. Singh, M. Gupta, R. Prasad, "Soil moisture retrieval over agricultural region through machine learning and sentinel 1 observations," Front. Remote Sens., 5: 1513620, 2025.

[15] K. Gojiya, B. Karangiya, S. Chavda, S. Gaadhe, D. Gojiya, H. Rank, "Soil moisture modelling using remote sensing and artificial neural networks: A Study of devbhumi dwarka region, gujarat, india," Int. J. Environ. Clim. Change, 14(1): 1-13, 2024.

[16] H. Karim Tabbahfar, F. Tabib Mahmoudi, "Optimum spectral indices for water bodies recognition based on genetic algorithm and sentinel-2 satellite images," J. Electr. Comput. Eng. Innovations, 12(1): 217-226, 2024.

[17] A. Singh, K. Gaurav, "Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images," Sci, Rep., 13(1): 2251, 2023.

[18] M. Lamichhane, S. Mehan, K. R. Mankin, "Soil moisture prediction using remote sensing and machine learning algorithms: A review on progress, challenges, and opportunities," Remote Sens., 17(14): 2397, 2025.

[19] K. Mao, H. Wang, J. Shi, E. Heggy, S. Wu, S. M. Bateni, "A general paradigm for retrieving soil moisture and surface temperature from passive microwave remote sensing data based on artificial intelligence," Remote sens., 15(7): 1793, 2023.

[20] J. Zeng, J. Peng, W. Zhao, C. Ma, H. Ma, "Microwave remote sensing of soil moisture," Remote Sens., 15(7): 4243, 2023.

[21] Y. Wang, J. Zhao, Z. Guo, H. Yang, N. Li, "Soil moisture inversion based on data augmentation method using multi-source remote sensing data," Remote Sens., 15(7): 1899, 2023.

[22] G. Wang, C. Wei, L. Yan, J. Li, "Soil moisture prediction model based on improved GRU recurrent neural network," Strategic Plann. Energy Environ., 43(2): 381-400, 2024.

[23] M. Baumberger, B. Haas, W. Tewes, B. Risse, N. Meyer, H. Meyer, "Gated recurrent units for modelling time series of soil temperature and moisture: An assessment of performance and process reflectivity," Environ Modell. Software, 183: 106245, 2025.

[24] W. Zheng, K. Zheng, L. Gao, L. Zhangzhong, R. Lan, L. Xu, "GRU–transformer: a novel hybrid model for predicting soil moisture content in root zones," Agronomy, 14(3): 432, 2024.

[25] M. R. K. Munaganuri, P. Gonnuri, R. Badiguntla, A. Kasukurthi, S. L. Sarma, S. M. Basha, "Deep learning-based crop prediction using LSTM and GRU for sustainable agriculture," Int. J. Environ. Sci., 11(11s): 538-545, 2025.

[26] M. Ahmed, R. C. Deo, N. Raj, A. Ghahramani, Q. Feng, Z. Yin, "Deep learning forecasts of soil moisture: convolutional neural network and gated recurrent unit models coupled with satellite-derived MODIS, observations and synoptic-scale climate index data," Remote Sen., 13(4): 554, 2021.

[27] A. Vyas, S. Bandyopadhyay, "Dynamic structure learning through graph neural network for forecasting soil moisture in precision agriculture," arXiv preprint arXiv:2012.03506, 2020.

[28] W. Wang, Y. Wei, L. Hao, Z. Wei, T. Zhao, "Soil moisture forecasting in wireless sensor networks via spatiotemporal graph convolutional networks," Vadose Zone J., 24(1): e70000, 2025.

[29] V. Grubišić, D. Vasić, H. Ljubić, R. Rozić, T. Volarić, "Soil moisture prediction with attention-enhanced models: a deep learning approach," Authorea Preprints, 2024.

[30] M. ElSaadani, E. Habib, A. M. Abdelhameed, M. Bayoumi, "Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations," Front. Artif. Intell., 4: 636234, 2021.

[31] J. Liu, F. Rahmani, K. Lawson, C. Shen, "A multiscale deep learning model for soil moisture integrating satellite and in situ data," Geophys. Res. Lett., 49(7): e2021GL096847, 2022.

[32] W. Dorigo, I. Himmelbauer, D. Aberer, L. Schremmer, I. Petrakovic, L. Zappa, "The International Soil Moisture Network:

serving Earth system science for over a decade," Hydrol. Earth Syst. Sci. Discuss., 25(11): 5749-5804, 2021.

[33] B. K. Iwana, S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," Plosone, 16(7): p. e0254841, 2021.

[34] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, "Time series data augmentation for deep learning: A survey," arXiv preprint arXiv:2002.12478, 2020.

[35] H. J. Diamond, T. R. Karl, M. A. Palecki, C. B. Baker, J. E. Bell, R. D. Leeper, "US Climate Reference Network after one decade of operations: Status and assessment," Bull. Am. Meteorol. Soc., 94(4): 485-498, 2013.

Biographies

Reyhaneh Bagheri received her Bachelor's degree in Surveying and mapping engineering from Isfahan University, Isfahan, Iran, in 2022. Since 2023, she is the M.Sc. student of remote sensing in Geomatics department of the faculty of Civil Engineering, Shahid Rajaee Teacher Training University.

- Email: reyhaneh.bagheri888@sru.ac.ir
- ORCID: 0009-0004-2267-9816
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Fatemeh Tabib Mahmoudi received her Bachelor's degree in Geomatics engineering (Surveying and mapping) from Khajeh Nasiredin Toosi University, Tehran, Iran, in 2004. She received his M.Sc. and PhD degrees in Photogrammetry from Tehran University, Tehran, Iran, in 2009 and 2014, respectively. Since 2016, she has been working as an assistant professor in the Geomatics department of the faculty of Civil Engineering, Shahid Rajaee Teacher Training University. She has some publications in the field of remote sensing data analysis, pattern recognition and data fusion.

- Email: fmahmoudi@sru.ac.ir
- ORCID: 0000-0002-8414-8189
- Web of Science Researcher ID: V-8360-2019
- Scopus Author ID: 36669646900
- Homepage: NA

AmirHossein Gholamian received his Bachelor's degree in Surveying and mapping engineering from National Geographic Organization University, Tehran, Iran, in 2021. Since 2022, he is the M.Sc. student of remote sensing in Geomatics department of the faculty of Civil Engineering, Shahid Rajaee Teacher Training University.

- Email: amirhoseinghloamian@sru.ac.ir
- ORCID: 0009-0004-8287-5480
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA