Applied Soft Computing Journal 193 (2026) 114845

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

SAGE-ET: Scalable and adaptive generalized estimation of ET_0

Gitika Sharma [a], Aayush Tripathi [a], Atharva I.P. Agrawal [a], Himanshu Sharma [b,*], Laxmi Kant Sharma [c]

[a] School of Computer Science and Engineering, Bennett University, Greater Noida, 201310, India
[b] Sharda School of Computer Science and Engineering, Sharda University, Greater Noida, 201310, Uttar Pradesh, India
[c] Department of Environmental Science, Central University of Rajasthan, Ajmer, 305817, Rajasthan, India

HIGHLIGHTS

- Proposes a novel federated learning (FL) framework for predicting reference evapotranspiration (ET_0).
- Integrates ConvLSTM architecture with FedProx optimization to ensure privacy preservation, scalability, and generalizability.
- Evaluates performance across four distinct climatic regions: arid, semi-arid, tropical wet and dry, and humid subtropical.
- Demonstrates superior performance over centralized ConvLSTM models in terms of R^2, MAE, and RMSE.
- Shows robustness to data heterogeneity and highlights the potential of FL in decentralized climate forecasting and sustainable water resource management.

ARTICLE INFO

Keywords:
Reference evapotranspiration
Limited climate data
Federated learning
FedProx
Generalized model

ABSTRACT

The estimation of reference evapotranspiration (ET_0) remains a challenging task, particularly in heterogeneously distributed data environments. Traditional physics-based models rely on localized climatic data, which limits their effectiveness in data-scarce regions. Similarly, centralized artificial intelligence models often face privacy risks and lack generalization when applied to heterogeneous environments and evolving environmental conditions. To address these challenges, we introduce SAGE-ET—a Scalable and Adaptive framework for the Generalized Estimation of ET_0. This model integrates an innovative federated learning architecture that combines ConvLSTM networks with a customized FedProx optimization strategy, termed FedET_0-Prox, resulting in a robust, privacy-aware, and scalable ET_0 forecasting framework applicable across diverse climatic zones. The proposed SAGE-ET framework is evaluated on climate datasets representing four distinct climatic regions—arid, semi-arid, tropical wet-dry, and humid subtropical—using two subsets of meteorological features. Experimental results confirm the superior performance of SAGE-ET compared to traditional centralized ConvLSTM models, demonstrating greater resilience to data distribution shifts, reduced root mean square error (RMSE), ranging from 0.0508 to 0.0310 mm/day, and an improved coefficient of determination (R^2), ranging from 0.9377 to 0.9744. This framework represents the next generation of distributed AI systems for environmental intelligence, highlighting the potential of federated learning to balance data privacy, collaborative intelligence, and scalability in complex, heterogeneous, and distributed contexts. SAGE-ET outperforms prior hybrid DL models (CNN-LSTM and TCN-BiLSTM) in RMSE and R^2 via federated generalization.

1. Introduction

Reference evapotranspiration (ET_0) is a fundamental measure in irrigation science, indicating the amount of water required by crops under standard environmental conditions. Accurate estimation of ET_0 guides irrigation planning and supports sustainable agriculture. This value represents the combined process of evaporation and plant transpiration from a reference surface. In practice, the FAO Penman–Monteith equation is widely adopted for ET_0 calculation, drawing on meteorological data such as air temperature, solar radiation, wind speed, and humidity [1]. However, in situations where comprehensive climate data collection is limited, conventional methods for estimating

* Corresponding author.
Email addresses: gitika.sharma@bennett.edu.in (G. Sharma), aayushtripathi09@gmail.com (A. Tripathi), agrawalip1601@gmail.com (A.I.P. Agrawal), er.himanshusharma92@gmail.com (H. Sharma), laxmikant_evs@curaj.ac.in (L.K. Sharma).

https://doi.org/10.1016/j.asoc.2026.114845
Received 24 June 2025; Received in revised form 14 January 2026; Accepted 10 February 2026
Available online 16 February 2026
1568-4946/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

reference evapotranspiration often yield less reliable results and may not accurately reflect local conditions [2,3].

Machine learning (ML) has recently emerged as a robust tool for overcoming these limitations. ML techniques, when applied to data collected from field sensors, enable more accurate assessment of crop water requirements and contribute to smarter, data-informed strategies for managing agricultural irrigation [2]. Additionally, innovations in satellite-based monitoring have further enhanced the timeliness and accuracy of ET_0 prediction [3,4]. However, most ML models operate optimally within the regions on which they were trained, limiting their broader applicability.

As agricultural activities spread across a wide range of environmental conditions, there is a growing demand for ET_0 prediction models that are broadly applicable and capable of adapting to diverse climatic regions [5,6]. Many existing methods rely heavily on parameters specific to local conditions, which often limits their accuracy and applicability when extended to different geographical regions [4,5]. Improving model adaptability contributes to more reliable irrigation strategies, helping mitigate the risks posed by climate variability. Research indicates that many currently available models face challenges with scalability and adaptability, especially when exposed to unfamiliar data [5]. Ensemble modeling, as proposed by Shi [7], leverages observations from multiple weather stations to improve generalization. Despite their effectiveness, centralized solutions often face practical limitations related to data access, which highlights the relevance of distributed modeling strategies.

Climate complexity adds more challenges to ET_0 modeling. Models trained on limited data often overfit and struggle with new conditions [6,8]. Predictive models often lose accuracy when the underlying data fail to capture the full spectrum of climatic variability present across different regions. Additionally, effective models must account for temporal trends, such as seasonal and interannual variations, which add another layer of complexity to accurate forecasting [8].

1.1. Decentralized modeling approaches

The above-mentioned issues highlight the need for a decentralized framework that addresses problems related to data heterogeneity while ensuring that the forecasting abilities of the model remain intact. Federated Learning (FL) emerges as a solution to mitigate this problem. Unlike centralized methods, FL enhances scalability and enables privacy by allowing multiple clients to collaboratively train a model without sharing their localized data. The ability of FL to handle varying and nonconforming data [9,10] can bring a positive change to ET_0 prediction by supporting generalized and reliable modeling across diverse regions.

Federated Learning is a machine learning framework that has revolutionized fields where privacy preservation and decentralization of model training across multiple clients is essential. Data privacy is assured in the federated framework by keeping client data localized and not transmitting raw data; instead, clients share model parameters to collaboratively train a common global model [9,11,12]. Federated Learning increases the adaptability of the model to real-world scenarios due to the enormous variety of input and knowledge it receives from these distinct regions and their climatic conditions [13].

1.2. Federated learning in agriculture and climate science

The adoption of FL in agricultural and environmental research is expanding. Recent studies demonstrate that collaborative learning among distributed data sources can enhance the reliability of forecasts, such as crop yield or water demand, while protecting sensitive local information [14]. Sensor networks using FL can efficiently gather and process climatic data, supporting timely decision-making in resource management [13–15]. Given the inherent diversity of climate data, federated approaches are particularly well-suited to addressing these challenges.

Building on earlier research which highlighted the strengths of ConvLSTM models in sparse data contexts [16], this study introduces the SAGE-ET model. By combining ConvLSTM with the FedProx optimization strategy, the new approach achieves greater scalability and maintains accuracy across varying climatic regions, addressing critical limitations of centralized learning frameworks.

1.3. Contributions of this study

In this work, a Federated Learning (FL) framework is proposed to predict reference evapotranspiration (ET_0) in the areas of different climatic conditions. The main contributions of this paper are:

- *Decentralized ET_0 estimation:* The present study proposes ConvLSTM-based Federated Learning (FL) framework—*SAGE-ET*, which takes advantage of a sequential deep learning architecture to capture time-dependent relationships in climate data, and uses decentralized environment to collaborate climate data of multiple stations while preserving data privacy.
- *Cross-regional generalization of et_0 prediction:* in this study, data from different climatic regions is employed for the generalized prediction of reference evapotranspiration. The risk due to overfitting to site-specific climatic conditions is also diminished by this approach, therefore qualifying for broader applicability.
- *FedProx integration for enhanced model robustness:* FedProx algorithm is integrated into the proposed FL framework to make model more robust and capable of making reliable ET_0 predictions, even in the presence of non-independently and identically distributed (non-IID) data.
- *Superiority over hybrid dl baselines:* The proposed SAGE-ET scheme outperforms centralized hybrid DL baselines (e.g., CNN-LSTM, TCN-BiLSTM) across heterogeneous climates through its novel FedProx- ConvLSTM integration, enabling privacy-preserving and generalized ET_0 prediction.

2. Background

Federated learning (FL) is a decentralized machine learning approach that enables multiple participants—ranging from mobile devices to organizations and edge computing nodes—to collaboratively train a shared global model, all while keeping their individual raw data private and local. This method uses decentralized computer resources, lowers communication costs, and addresses concerns about privacy. Since data is sensitive, geographically scattered, or subject to legal restrictions, FL is extremely beneficial (e.g., financial, medical, and IoT systems).

The most common algorithm in FL is Federated Averaging (FedAvg) [17], which consists of regular model update aggregating on a central server and local training on clients. The following is the mathematical definition of FedAvg. Every client j updates parameters of its local model θ_j by minimizing the local loss function $G_j(\theta)$ with the help of stochastic gradient descent (SGD):

$$\theta_j^{r+1} = \theta_j^r - \alpha \nabla G_j(\theta_j^r), \tag{1}$$

here α is the learning rate and r is the iteration index. The global model is subsequently generated by the server by combining the local updates. θ^{r+1}:

$$\theta^{r+1} = \sum_{j=1}^{M} \frac{m_j}{m} \theta_j^{r+1}, \tag{2}$$

where m_j is the number of data samples on client j, m is the total number of samples across all clients, and M is the total number of clients.

FedAvg has become popular, although it has several limitations. First, poor convergence and model performance could originate from statistical heterogeneity between clients caused by non-IID (non-independent and identically distributed) data [18]. Second, system heterogeneity, where the process of aggregation could be slowed by stragglers, such as variations in client network conditions and computational capacity [19]. Third, communication overhead from the regular

connection of client-server relationships can be expensive, especially in settings with limited resources [20].

To address these challenges, Federated Proximal (FedProx) [21] was proposed as an extension of FedAvg. FedProx introduces a proximal term to the local objective function to stabilize training and handle heterogeneity. The local optimization problem in FedProx is formulated as:

$$\min_{\theta} G_j(\theta) + \frac{\lambda}{2}\|\theta - \theta^r\|^2, \tag{3}$$

where λ is a hyperparameter controlling the impact of the proximal term, and θ^r is the global model at iteration r. This term ensures that local updates do not deviate significantly from the global model, improving convergence in heterogeneous settings.

Even though FedProx overcomes certain drawbacks of FedAvg, it introduces new drawbacks. Indicatively, the value of λ selected in FedProx can be of great significance to performance and requires special adjustment [22]. Moreover, an increase in the number of clients causes FedAvg and FedProx to have scalability problems, which intensifies the computing and communication expenses [20]. In addition, although FL safeguards data privacy by default, model updates may still disclose private information, demanding the use of additional privacy-preserving techniques like secure aggregation or differential privacy [23,24].

To sum everything up, federated learning is a feasible method for decentralized machine learning that safeguards privacy. To achieve its full potential, however, difficulties with scalability, heterogeneity, and communication effectiveness must be resolved. Although new innovations such as FedProx and adaptive optimization techniques are positive, further research is required to make FL trustworthy and useful for real-life applications.

2.1. Hybrid deep learning

Prior hybrid deep learning models for ET_0 prediction, such as CNN-LSTM and TCN-BiLSTM, have used convolutional and recurrent layers to capture spatiotemporal climate dependencies. However, these approaches suffer from overfitting in cross-regional tests due to centralized training, poor generalization under non-IID data distributions common in heterogeneous agricultural regions, and vulnerability to privacy risks when handling distributed farm-level data [25,26].

SAGE-ET addresses these gaps by pioneering the integration of ConvLSTM within a FedProx-based federated learning framework. This enables privacy-preserving, scalable ET_0 prediction across non-IID climates (from arid to humid subtropical) without client drift, overcoming the limitations of centralized hybrids and prior works that assumed IID data [27,28].

3. Problem formulation

3.1. Objective

This research focuses on improving convergence guarantees for FedProx in a federated framework marked by statistical heterogeneity, partial device participation, and inexact local updates [12]. We analyze FedProx's performance in ET_0 generalization. Furthermore, we explore solver-agnostic adaptations of FedProx to support flexible local solvers across heterogeneous climate data sources. By addressing these challenges, this research seeks to improve the generalization and the practical implementation of federated learning models for climate models as well as agricultural decision-making.

3.2. Mathematical formulation

Consider a federated learning system having N clients, each client i possessing a local dataset $\mathcal{D}_i$. The global optimization objective will be as follows:

$$\min_{\theta} J(\theta) = \frac{1}{N}\sum_{i=1}^{N} J_i(\theta) \tag{4}$$

Here, θ represents the global model parameters, $J_i(\theta)$ denotes the local loss function at client i, and this is defined as:

$$J_i(\theta) = \mathbb{E}_{(X,y)\sim D_i}[F_i(\theta; X, y)]. \tag{5}$$

The FedProx framework introduces a proximal term [12]:

$$\min_{\theta} G_i(\theta; \theta^r) = J_i(\theta) + \frac{\lambda}{2}\|\theta - \theta^r\|^2 \tag{6}$$

where θ^r is the global model at global round r and λ is the proximal term coefficient controlling the regularization strength. The local update term $\frac{\lambda}{2}\|\theta - \theta^r\|^2$ ensures that local updates do not diverge significantly but remain close to the global model, even under highly diverse circumstances [29,30].

At every global round r, the server selects a group of clients C_r to participate in the training. The local training of each chosen client $i \in C_r$ is performed by minimizing the surrogate objective $G_i(\theta; \theta^r)$. After local training for E epochs, all participating clients transmit their updated models $\theta_i^{(r+1)}$ to the server. The server then aggregates these local models to form the new global model:

$$\theta^{r+1} = \frac{1}{|C_r|}\sum_{i\in C_r} \theta_i^{r+1} \tag{7}$$

where $|C_r|$ represents the number of clients participating in the current round.

The FedProx framework incorporates several key mathematical components that clarify its formulation. The local objective function $J_i(\theta)$ quantifies the model performance for client i, while the global model weights θ^r serve as the reference for local updates. The surrogate objective function $G_i(\theta; \theta^r)$ balances both local updates and global stability by combining the local loss and proximal regularization terms. The proximal coefficient λ ensures smoothness of the local updates, preventing them from drifting significantly away from the global model, thereby maintaining training stability. The local update $\theta_i^{(r+1)}$ represents the adjusted weights of client i obtained after solving the local optimization problem over E epochs. By leveraging these mathematical principles, FedProx achieves greater stability and efficiency in federated learning for ET_0 estimation compared to FedAvg [12,30].

4. Methodology

The flow of execution of the model is shown in Fig. 1. The figure illustrates three primary sections: the federated learning framework for predicting ET_0, local training for each region as clients, and the updating of the global model through the consolidation of local model updates.

This section highlights the proposed SAGE-ET framework - a Federated Learning based solution for predicting ET_0. The federated model is designed with four regions, also referred to as clients, where each client's local model trains on its local data. Updates from local models are aggregated using the FedProx algorithm, which ensures that ET_0 predictions are generalized across all regions. Though other federated optimization algorithms such as FedAvg [17] and FedSGD [18] are well-established, FedProx [21] is chosen in this research due to its theoretical adaptability to heterogeneous and non-IID data. Client drift is a known issue in FedAvg within federated settings where data distributions among clients are highly heterogeneous, as is the case in our multi-regional climatic stations, leading to slower convergence and poorer generalization. Although FedSGD is mathematically stable, it incurs higher communication costs and is less feasible for large-scale or bandwidth-constrained applications. FedProx introduces a proximal regularization term into each client's local objective, constraining local updates to remain closer to the global model and thereby reducing client divergence. As observed in previous studies [12,21,30], FedProx achieves more stable and consistent convergence across heterogeneous networks, making it a more suitable choice for this feasibility study. The proposed SAGE-ET framework is designed to handle heterogeneous

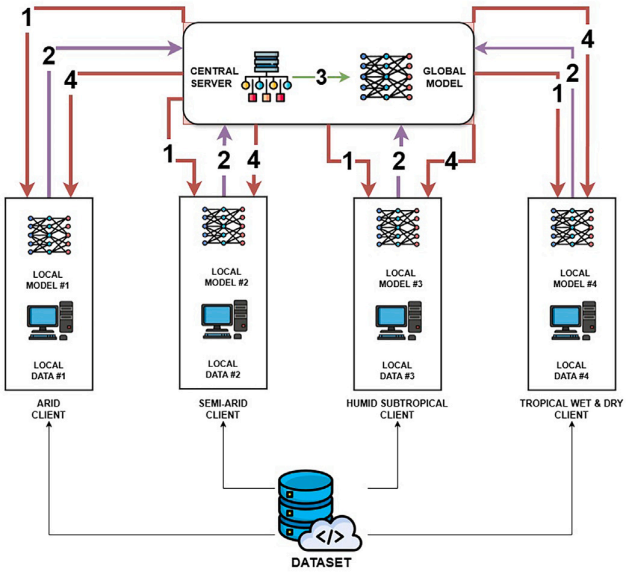


Fig. 1. Flow of execution for the federated learning based SAGE-ET framework. **The execution flow, as depicted in Fig. 1, follows these annotated steps: • Arrow 1 (Server → Clients):** Initial global model is sent to all clients. • **Arrow 2 (Clients → Server):** Clients train locally with FedProx and upload weight updates to the server. • **Arrow 3 (Server Aggregation):** Server computes new global model. • **Arrow 4 (Server → Clients):** Server shares the improved global model with clients for the next round.

climate data conditions by leveraging the FedProx optimization mechanism, which makes it adaptive to different local data distributions. Consequently, rather than conducting experimental comparisons at this stage, the selection of FedProx is conceptually motivated—based on its theoretical advantages and alignment with the heterogeneous, multi-climatic data scenario addressed in SAGE-ET. Future extensions of this work will benchmark other federated optimizers (e.g., FedAvg, FedSGD, and SCAFFOLD) under the same experimental setup to quantitatively assess their performance. Section 4.2 explores the federated structure in detail, including the application of the ConvLSTM-based model and the FedProx algorithm.

Because of the problem of data scarcity that arises in some areas, the use of a minimalistic method to predict ET_0 accurately is needed. To address this, ET_0 predictions are also evaluated in the study using three and five parameters. These input feature sets have been chosen depending on the findings of our previous studies [16,31], where T_{max} (maximum temperature), T_{min} (minimum temperature), and R_s (solar radiation) were found to be important climatic parameters for reliable ET_0 prediction. Furthermore, our former research [16] discloses that the inclusion of relative humidity (R_h) and wind speed (U_2) resulted in outputs comparable to the FAO-56 method.

This feature selection methodology, based on prior interpretability assessments, provides a natural degree of explainability by ensuring that the selected features are scientifically significant and empirically verified. Hence, inspired by these observations based on our previous work, SAGE-ET has been tested with minimal (3-feature) and comprehensive (5-feature) input datasets to include both data-scarce and data-rich scenarios.

4.1. Experiment area and data description

The study is conducted in four different climatic regions of India, each region having two meteorological stations. The regions are:

Hot Desert Climate (Arid), with two stations, Barmer and Jodhpur, located in Rajasthan. Barmer is located at Latitude 25.7523 and Longitude 71.4102, with an annual rainfall of about 285 mm and an elevation of 209.15 meters, while Jodhpur is located at Latitude 26.2722 and Longitude 73.0219, with an annual rainfall of about 340 mm and an elevation of 257.67 meters. This region experiences very hot summers and mild winters along with low rainfall.

Northwestern Indian Climate (Semi-Arid), with two stations, Ludhiana, located in Punjab, and New Delhi. Ludhiana is located at Latitude 30.9156 and Longitude 75.8518, with an annual rainfall of about 655 mm and an elevation of 240.26 meters, while New Delhi is located at Latitude 28.6361 and Longitude 77.1959, with an annual rainfall of about 760 mm and an elevation of 206.37 meters. This region experiences hot summers and cold winters, with temperatures ranging from above 40 °C in summer to around 5 °C in winter, along with moderate rainfall.

Savanna Climate (Tropical Wet and Dry), with two stations, Coimbatore and Thanjavur, located in Tamil Nadu. Coimbatore is located at Latitude 10.9962 and Longitude 76.9646, with an annual rainfall of about 1150 mm and an elevation of 540.94 meters, while Thanjavur is located at Latitude 10.7755 and Longitude 79.1307, with an annual rainfall of about 1050 mm and an elevation of 39.16 meters. This region experiences tropical dry and wet weather, with a good amount of rainfall, hot and humid summers, and mild winters.

Northeastern Indian Climate (Humid Subtropical), with two stations, Guwahati and Dibrugarh, located in Assam. Guwahati is located at a Latitude of 26.1532 and a Longitude of 91.7638, with an annual rainfall of about 2000 mm and an elevation of 400.01 meters, while Dibrugarh is located at a Latitude of 27.4819 and a Longitude of 94.9204, with an annual rainfall of about 2400 mm and an elevation of 182.3 meters. This region experiences subtropical humid weather, with a high amount of rainfall, hot and humid summers, and mild winters, along with persistent cloud cover and distinct seasonal variations, all of which majorly impact the precise measuring of solar radiation. Studies have indicated that the calibration of solar radiation is remarkably tricky due to increased diffuse radiation, which is caused by recurring cloud covers and moisture [32].

The datasets were grouped based on climatic area: data of the stations within the same climatic area (e.g., Barmer and Jodhpur in the case of arid) were pooled into one client representing the same climatic region. Every client is, therefore, identified by a unique climate area, rather than a single station. These stations recorded the climatic data for a period of 21 years, from 2000 to 2021. Statistical analysis for each region's data is presented in Table 1. Fig. 2 shows heatmaps of cross-correlating meteorological parameters and ET_0 for each region.

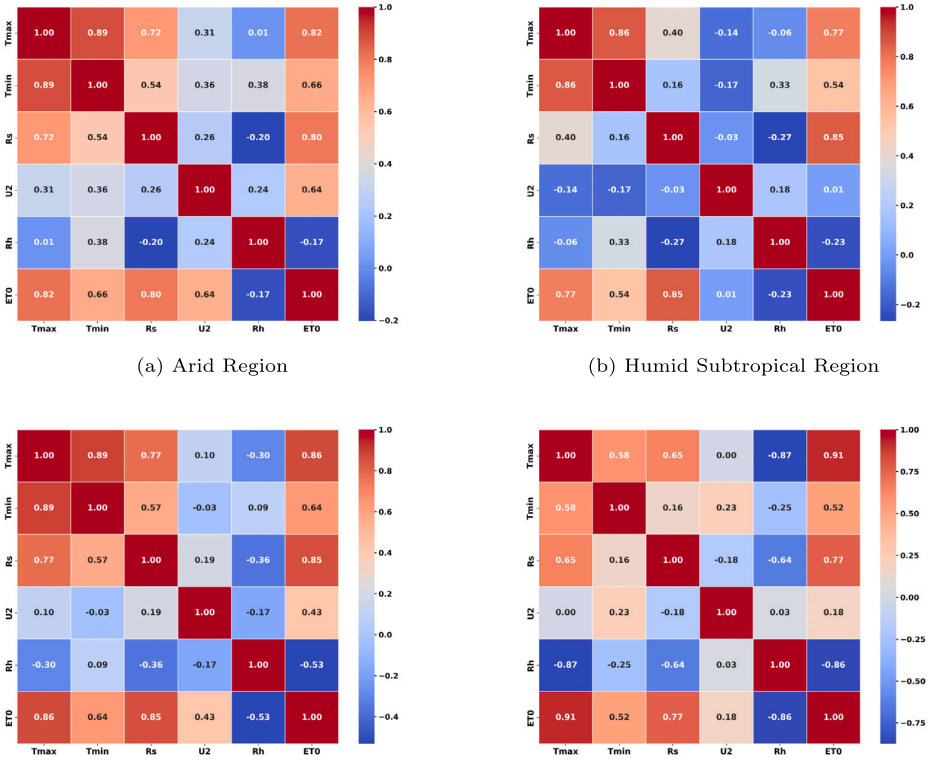
To serve as a benchmark dataset for evaluating model performance, ET_0 values were calculated using the FAO-56 Penman-Monteith equation. Climatic inputs such as minimum temperature (T_{min}), maximum temperature (T_{max}), wind speed (U_2), relative humidity (R_h), solar radiation (R_s), and sunshine hours (SSH), which are used for predictions, were collected for each region from the India Meteorological Department (IMD), Pune, India.

4.2. Execution flow

Our research presents an FL-based SAGE-ET framework utilizing the FedProx algorithm to improve the consistency and precision of the ET_0 estimation and extract information from meteorological data in four distinct climatic regions of India. At the heart of our approach is a Convolutional LSTM model, trained within a federated setup using the FedProx algorithm, to capture the spatial and temporal dependencies inherent in climate data. These patterns are necessary for accurate ET_0 estimation across distinct climatic conditions. The climatic datasets used in this research are multivariate temporal sequences without predefined spatial grids. Since ConvLSTM layers require a five-dimensional input format (samples, timesteps, height, width, channels), the study mapped the climatic features to the channel dimension and fixed the spatial dimensions to 1×1. This resulted in each input instance being defined as

Table 1
Statistical observations of the meteorological data for the four regions.

Training dataset								Testing dataset						
M.D.	X_{min}	X_{max}	X_{mean}	X_{std}	X_{skew}	X_{kurt}	C.V.	X_{min}	X_{max}	X_{mean}	X_{std}	X_{skew}	X_{kurt}	C.V.
Arid region														
T_{max}	15.73	47.65	34.37	6.02	−0.39	−0.75	0.18	19.47	47.23	34.53	5.93	−0.41	−0.73	0.17
T_{min}	1.67	34.10	20.22	7.28	−0.46	−1.03	0.36	2.66	33.55	20.32	7.18	−0.51	−0.93	0.35
R_s	2.09	29.95	20.28	4.60	−0.29	−0.43	0.23	2.99	29.34	20.44	4.51	−0.31	−0.36	0.22
U_2	0.49	9.15	2.83	1.21	1.13	1.22	0.43	0.46	9.08	2.83	1.22	1.12	1.24	0.43
R_h	4.38	93.38	37.49	17.57	0.54	−0.47	0.47	5.00	88.00	37.05	17.59	0.56	−0.52	0.47
ET_0	1.31	14.35	6.51	2.41	0.69	−0.30	0.37	1.67	14.51	6.56	2.38	0.70	−0.22	0.36
Humid subtropical region														
T_{max}	15.87	43.27	29.47	4.34	−0.31	−0.64	0.15	16.15	42.92	29.56	4.34	−0.31	−0.58	0.15
T_{min}	2.78	29.53	19.19	5.59	−0.41	−1.06	0.29	4.12	29.03	19.26	5.52	−0.45	−1.01	0.29
R_s	0.65	28.98	14.94	4.78	−0.06	−0.28	0.32	2.12	28.08	15.12	4.83	−0.09	−0.30	0.32
U_2	0.31	4.13	1.17	0.46	1.75	4.93	0.39	0.30	3.76	1.16	0.45	1.61	3.95	0.39
R_h	29.69	95.94	74.01	10.66	−0.71	0.08	0.14	29.75	95.94	74.04	10.63	−0.67	0.11	0.14
ET_0	0.74	7.81	3.15	1.11	0.51	−0.35	0.35	0.82	8.76	3.18	1.12	0.44	−0.35	0.35
Semi-arid region														
T_{max}	11.17	49.99	32.87	7.40	−0.15	−0.81	0.23	13.83	48.59	32.95	7.30	−0.14	−0.80	0.22
T_{min}	−0.97	35.30	19.08	8.19	−0.26	−1.19	0.43	−1.36	34.38	19.18	8.13	−0.30	−1.13	0.42
R_s	1.48	30.02	17.76	5.58	−0.21	−0.63	0.31	2.63	29.63	17.82	5.51	−0.17	−0.65	0.31
U_2	0.41	6.98	2.06	0.76	0.96	1.36	0.37	0.52	5.83	2.04	0.74	0.99	1.53	0.36
R_h	4.19	95.12	44.41	19.56	0.33	−0.70	0.44	5.56	93.06	44.43	19.56	0.33	−0.69	0.44
ET_0	0.68	15.40	5.23	2.37	0.80	0.25	0.45	0.96	14.14	5.22	2.34	0.85	0.44	0.45
Tropical wet and dry region														
T_{max}	21.59	42.97	31.80	3.92	0.26	−1.02	0.12	21.78	42.44	31.70	3.90	0.30	−0.93	0.12
T_{min}	11.93	29.88	22.24	3.11	−0.30	−0.14	0.14	12.46	29.65	22.20	3.08	−0.35	−0.02	0.14
R_s	0.76	27.25	18.65	4.14	−0.88	0.73	0.22	0.54	26.42	18.57	4.26	−0.87	0.66	0.23
U_2	0.44	7.41	2.74	1.05	0.61	0.35	0.38	0.40	6.93	2.74	1.03	0.56	0.27	0.38
R_h	26.75	95.44	73.08	12.66	−0.43	−0.64	0.17	30.06	95.31	73.47	12.44	−0.47	−0.64	0.17
ET_0	0.98	10.00	4.63	1.54	0.48	−0.37	0.33	0.86	9.76	4.58	1.52	0.49	−0.32	0.33



(a) Arid Region (b) Humid Subtropical Region

(c) Semi-Arid Region (d) Tropical Wet and Dry Region

Fig. 2. Heatmaps of cross-correlating meteorological parameters and ET_0 for each region.

(samples, timesteps, 1, 1, features), where the "features" represent the meteorological variables (T_{max}, T_{min}, R_s, U_2, and R_h). This setup allows the ConvLSTM to extend its spatiotemporal learning capabilities to multivariate temporal sequences by convolving across feature channels and capturing temporal dependencies across successive timesteps.

Before training, comprehensive preprocessing steps are taken to ensure data consistency, model robustness, and reliability. Normalization is executed using the MinMaxScaler approach, scaling each meteorological parameter within a (0,1) range. Normalization is essential in preventing bias from dominant features [33]. Outlier elimination is done using the Interquartile Range (IQR) method, a well-recognized and established method to effectively identify and then eliminate extreme values that could negatively impact the model performance [34]. This step ensures that there are no anomalies in the dataset that could negatively affect the training process. In the SAGE-ET architecture, normalization is performed at each client based on locally computed minimum and maximum values, thereby avoiding the use of centralized statistics. This design aligns with the privacy-conscious philosophy of federated learning by eliminating the flow of raw data metrics and reducing the coordination overhead at the central server. The computational feasibility and empirical validity of such client-side normalization have been demonstrated in the federated learning literature [19,21], where clients independently preprocess their data before local training, focusing on managing heterogeneity algorithmically rather than through shared global input scaling. Lastly, the data is divided into training and testing sets, with a portion of the combined training data being pretrained at first on the central model. Pre-training is a well-known and common practice within the federated learning paradigm to improve convergence in distributed systems [35].

The latter is also supported by recent research, namely, Yu et al. [36] point out that pre-training results in greater model robustness and reduction in communication overhead in federated environments of heterogeneous nature. Also, Wang et al. [37] highlight its applicability in enhancing the performance and stability of federated models, especially when it comes to non-IID (non-independent and identically distributed) data.

The reason why these preprocessing techniques are being used is justified by their efficiency in improving model performance and preserving data quality. Normalization mitigates feature scaling discrepancies, the IQR method improves data consistency by systematically eliminating outliers, and pre-training facilitates a robust initialization framework for the federated learning process, collectively contributing to the development of a robust, accurate, and reliable ET_0 estimation model. To guarantee the model's outputs apply in any climate region, top hyperparameter values were selected and optimized with the Optuna framework, a powerful automatic hyperparameter tuning tool. Because of this, trials of different settings for the architecture and training parameters were possible, which greatly improved the results. There were 50 separate trials in the tuning stage, during which each trial trained a candidate model for 100 epochs on the entire data set. We used the validation loss as the primary objective for our optimization process.

The filters within the ConvLSTM layers were tested between 64 and 256, while kernel sizes were reduced from 5 to 2 to detect how things vary across the spatial domain. Also, reducing and increasing hidden state dimensionality between 64 and 256 and the number of stacked ConvLSTM layers between 2 and 5 was tested to allow the model to discover deeper spatiotemporal features. Between 0.1 and 0.4 dropout rates were checked and several common non-linear activation functions (ReLU, ELU, tanh) were used. Optimizing learning rates occurred using a logarithmic scale from 1×10^{-4} to 1×10^{-2}, and experiments were performed using batch sizes ranging from 64 to 256. An optimal configuration of the model was achieved using Optuna-based hyperparameter optimization. The final architecture of the model framework includes 192 convolution filters with a 5 kernel size and 128 hidden state dimensions. Further, five ConvLSTM layers with ReLU as activation function was used. A dropout rate of 0.30 was applied to avoid the issue of overfitting, and a learning rate of 0.00118 and a batch size of 224 were used. This configuration setting was used to train both the local and centralized models. The models were trained and implemented using TensorFlow/Keras on a CUDA-based system with GPU acceleration. All datasets were divided into **training (72 %)**, **validation (8 %)**, and **testing (20 %)** subsets through a temporal (chronological) partition. Specifically, the initial data samples were used for training, followed by validation, and the most recent samples were reserved for testing, corresponding to the years **2000–2018** for training/validation and **2019–2021** for testing. This approach ensured that the model was evaluated on unseen future data, thereby preventing information leakage across time. To ensure robust convergence and no overfitting, regularization techniques like a 0.30 dropout rate to ConvLSTM layers, using Min–Max feature normalizing for all features, and an early stopping (*monitor* = validation loss, *patience* = 5) to auto-restore the lowest validation error checkpoint were used during training. Table 2 shows the hyperparameter tuning search space and other criteria used.

Table 2
Hyperparameter tuning search space, selection criterion, and final configuration used.

Hyperparameter	Search space	Final value
Conv filters	64–256 (step 32)	192
Kernel size	2–5	5
Hidden dimension	64–256 (step 32)	128
Number of layers	2–5	5
Dropout rate	0.1–0.4 (step 0.05)	0.30
Activation fn	{ReLU, ELU, Tanh}	ReLU
Learning rate	$[1\times10^{-4}, 1\times10^{-2}]$ (log-uniform)	0.001178
Batch size	64–256 (step 32)	224
Epochs (fed setup)	Fixed at 100 (after tuning stage)	100

In a federated framework, the FedProx algorithm was integrated to ensure proper stability in heterogeneous client environments. Although most of the hyperparameters were sourced from the main model, a few were configured manually before federated execution. For every client, the local training continued for 100 epochs per round, while federated learning continued for 50 full global rounds. It provided an equal and repeatable way to train in a distributed system. Every client in the SAGE-ET model was trained for up to 100 local epochs in each communication round, with an early-stopping criterion (*patience* = 5) based on minimal validation loss, thereby promoting effective local convergence. This setup aligns with standard federated learning practices, where heterogeneity in local data distributions is managed through 5 to 100 local training epochs [21,38]. To ensure stable convergence, both local training and validation losses were consistently monitored on each client before transmitting model updates for global aggregation. Despite the fact that the global convergence paths were not drawn directly, the general assessment of the aggregated model through the application of different statistical and hydrological quantities—the Nash–Sutcliffe Efficiency (NSE), coefficient of determination (R^2), root mean square error (RMSE), and Kling–Gupta Efficiency (KGE)—exhibited successful convergence and stability even in the case of non-IID data distributions.

Though a majority of the hyperparameters were shared with the main model, some of them had to be configured manually and then executed in the federated setting. For instance, the proximal term coefficient (μ) was manually adjusted to 0.2 after some trials, which demonstrated that it was able to stabilize training by restricting the drift between client updates and the global model in our non-homogeneous data setting. Even though the value of the proximal coefficient μ = 0.2 was selected after initial exploratory stability tests, these internal evaluations were intended solely to ensure consistent convergence behaviour among clients, rather than to exhaustively optimize the hyperparameter. Specifically, the global loss trend and regularity of client updates were monitored for representative μ values (0.1, 0.2, and 0.3), with μ = 0.2 yielding the most stable loss reduction without

oscillations or divergence. This observation lies within the commonly reported stability range (0.1–0.5) in prior FedProx literature addressing heterogeneous environments [21,39], thereby conceptually supporting the choice without requiring extensive ablation curves. Since SAGE-ET serves primarily as a proof-of-feasibility, this selection was guided by theoretical reasoning and qualitative stability. Future work will extend this by systematically analyzing the sensitivity of μ and related federated optimization parameters through ablation studies and quantitative convergence analysis.

In the early stages, every client used 20% of its own local data to do a small pretraining step. This pretraining step was done entirely on the client side, and only model parameters that were generated as a result were sent to the central server, which then aggregates these weights from all the clients to initialize the global model. Thereafter, the standard federated learning process was applied. Importantly, no raw data were ever shared with the global model, ensuring that basic privacy compliance is followed by the framework. Thanks to this step, Separation and Orthogonality-based INitialization (SOIN) achieved faster learning and a well-initialized state when using non-IID data sets. Collectively, combining Optuna with FedProx federated optimization made it possible to achieve stable results and strong performance in all climates, despite the challenges of heterogeneous data and restricted communication.

The federated learning process in our study is orchestrated by the FedProx algorithm, which is an extension of the Federated Averaging (FedAvg) approach, particularly designed to mitigate challenges arising from non-IID across client data [38]. Federated Averaging (FedAvg) is a cornerstone algorithm in federated learning, where a central server iteratively aggregates updated weights from local models trained on decentralized client data, enhancing the performance of the global model with each iteration. Although effective, one of the main assumptions made by FedAvg is that data distributions of clients are IID, not much to be found in practice. In cases where the data in a client has non-IID characteristics, and local models are trained on this heterogeneous data, data distributions may become very divergent, causing degradation in the overall performance and convergence. This effect is typically termed client drift of the global model [40].

FedProx overcomes client drift, by using a proximal term in the local training objective function [12]. This is an added regularizer term, incorporated into the local objective function as $F_i(\theta_i^{r,e}; X, y) + \frac{\lambda}{2}||\theta_i^{r,e} - \theta^r||^2$, which penalizes local model changes that are substantially different from the existing global model weights, thus promoting better consistency of client models and speeding up convergence toward a more consistent and universal global model. The proximal coefficient, λ, controls the intensity of regularization and requires optimal performance tuning, along with the learning rate α. The FedProx algorithm operates in the following way, and its sequential protocol is outlined below:

The server initializes the global ConvLSTM model weights (θ^0), pretrained on a subset of the aggregated client data, and then distributes them to a random subset C_r of N total participating clients at each global iteration r. During the local training phase, each client i in C_r trains the received model on its own dataset for E local epochs, incorporating the FedProx proximal term into the loss function. After the local training phase, the updated model weights (θ_i^{r+1}) are sent back to the central server, which aggregates them to update the global model. The FedProx (and FedAvg) currently apply a standard aggregation technique where local model weights are averaged simply as $\theta^{r+1} = \frac{1}{|C_r|}\sum_{i\in C_r}\theta_i^{r+1}$ for all i in C_r, where $|C_r|$ denotes the number of clients in round r. This repetition of model distribution, local training, and weight aggregation happens until R global rounds are completed. The global model gradually integrates the statistical differences that exist among all the clients involved. In each iteration, the update to the global model parameters (θ^{r+1}) is communicated to the server, eventually producing the ultimate global model weights (θ^R) after completing the R global iterations of federated training.

Algorithm 1 FedProx framework.

Require: Number of global rounds R, proximal coefficient λ, number of clients N, initial global model weights θ^0, local epochs E, batch size B

for $r = 0$ to $R - 1$ **do**
  Server selects a random subset C_r of N clients
  Server sends global model θ^r to all selected clients in C_r
  for each client $i \in C_r$ **in parallel do**
    Initialize local model $\theta_i^{r,0} = \theta^r$
    for local epoch $e = 0$ to $E - 1$ **do**
      for each batch (X, y) of size B **do**
        Compute gradient: $g_i = \nabla F_i(\theta_i^{r,e}; X, y)$
        Update local model: $\theta_i^{r,e+1} \leftarrow \theta_i^{r,e} - \alpha(g_i + \lambda(\theta_i^{r,e} - \theta^r))$
      end for
    end for
    Set local update $\theta_i^{r+1} \leftarrow \theta_i^{r,E}$
    Send θ_i^{r+1} to the server
  end for
  Server aggregates local models:

$$\theta^{r+1} \leftarrow \frac{1}{|C_r|}\sum_{i\in C_r}\theta_i^{r+1}$$

  Save updated global model θ^{r+1}
end for
Output: Final global model θ^R

4.3. Evaluation of model parameters

A crucial step in model development is gauging its performance. The accuracy of the proposed model in replicating actual outputs is evaluated by comparing its predictions against observed values using established mathematical metrics. In this study, a federated framework is used with various climatic parameters to estimate the ET_0. A comparison between the forecasted ET_0 values and the actual/observed ET_0 values, which were computed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2), was performed. These statistical benchmarks are calculated using the following equations:

$$MAE = \frac{\sum_{i=1}^{N}\left(ET_{0\mathrm{Ti}} - ET_{0\mathrm{Pi}}\right)}{N} \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(ET_{0\mathrm{Ti}} - ET_{0\mathrm{Pi}}\right)^2}{N}} \tag{9}$$

$$R^2 = \frac{\sum_{i=1}^{N}\left(ET_{0\mathrm{Ti}} - \overline{ET_{0\mathrm{T}}}\right)\left(ET_{0\mathrm{Pi}} - \overline{ET_{0\mathrm{P}}}\right)}{\sum_{i=1}^{N}\left(ET_{0\mathrm{Ti}} - \overline{ET_{0\mathrm{T}}}\right)^2 \sum_{i=1}^{N}\left(ET_{0\mathrm{Pi}} - \overline{ET_{0\mathrm{P}}}\right)^2} \tag{10}$$

Here, the FAO-PM calculated value or observed ET_0 value is denoted by $ET_{0\mathrm{Ti}}$ ($mm\,d^{-1}$), $ET_{0\mathrm{Pi}}$ is the forecasted ET_0 value for the i^{th} data point computed by the model, the mean of the observed ET_0 values is represented by $\overline{ET_{0\mathrm{T}}}$, and the mean of the forecasted ET_0 values is represented by $\overline{ET_{0\mathrm{P}}}$. N is the total number of data points in the study.

Mean Absolute Error (MAE) varies between 0 and ∞ and represents the average difference between actual and forecasted values, ignoring direction. The lower the MAE value, the better the prediction made by the model [41]. Root Mean Square Error (RMSE) indicates the deviation in residual errors, showcasing the closeness of the data to the best-fit line. The lower the RMSE value, the better the prediction performed by the model. The correlation between forecasted and observed ET_0 values is highlighted by the Coefficient of Determination (R^2).

4.4. Diebold–mariano (DM) test

The comparison of predictive performance measures, including MAE, RMSE, and R^2, provides quantitative information on the accuracy of the model but does not offer any information on whether there are statistically significant differences between models. The Diebold–Mariano (DM) test was used to assess the significance of the predictive differences between the centralized ConvLSTM and federated SAGE–ET models. This is a statistical test that allows an official statistical analysis on determining whether or not two models used to give a prediction have the same predictive power within a specified time frame [42].

Where $ET_{0,obs,n}$ is the observed reference evapotranspiration and $ET^{(1)}_{0,pn}$, $ET^{(2)}_{0,pn}$ are the respective predicted values of two rival models. The forecast errors of the models are given as:

$$E^{(1)} = ET_{0,obs,n} - ET^{(1)}_{0,pn}, \quad E^{(2)} = ET_{0,obs,n} - ET^{(2)}_{0,pn}.$$

In order to measure predictive performance, the DM framework requires a loss function $L(\cdot)$ that measures the forecast accuracy. Two popular loss formulations are:

$$L(E) = \frac{1}{n}\sum |E_i| \quad \text{(absolute error loss)},$$

$$L(E) = \frac{1}{n}\sum E_i^2 \quad \text{(squared error loss)}.$$

These are two major forms of prediction error: the absolute loss evaluates the overall scale of deviation equally, and the squared loss evaluates the overall scale more severely; thus, the latter is more sensitive to outliers. With climatic time-series data, in which extreme weather factors (heatwave or sudden rain anomalies, etc.) may cause large yet infrequent errors, the absolute error formulation is more robust and interpretable for overall model performance. The final statistical comparison was therefore taken in the form of the absolute loss form, because it was stable and there was a physical interpretation in terms of daily variability of ET [43].

The DM test compares the null hypothesis of equal accuracy of the forecasts of the two models with the alternative hypothesis of different predictive accuracy [42]:

$$H_0 : E[d_t] = 0 \tag{11}$$

$$H_1 : E[d_t] \neq 0 \tag{12}$$

where $d_t = L(E^{(1)}_t) - L(E^{(2)}_t)$ is the difference in loss between the two models at time t. The DM statistic is as follows:

$$DM = \frac{\bar{d}}{\sqrt{LRV(d)/n}} \tag{13}$$

in which $\bar{d}$ denotes the sample mean of the differential loss series and LRV(d) is the long-run variance of the series of losses. The *Newey–West heteroskedasticity and autocorrelation consistent (HAC)* estimator [44] was used to estimate the long-run variance, accounting for autocorrelation and heteroskedasticity in the loss differentials. Time-dependent variables like daily evapotranspiration, where errors at any time usually have serially correlated successive errors, and heteroskedastic errors with seasonality, should be corrected in this way. Overlooking these dependencies could result in variance underestimation and inflated statistical significance. The Newey–West HAC method uses lag-weighted autocovariances, which give an unbiased and consistent standard error estimation even in the presence of serially correlated and non-homoskedastic data [43].

In order to make it even more robust in the case of non-IID data, the subsampling procedure was used, based on the refined DM methodology suggested by H. Chen [45]. In each regional dataset, 100 random subsamples of 300 observations were selected, and the DM statistic and the corresponding two-sided p-value were calculated on each subset. The average of these values was provided in order to have a stable inference. The null hypothesis is that the predictive accuracy is equal, and, under the null hypothesis, the DM statistic takes the form of a standard normal distribution in the limit. The null hypothesis is rejected when the statistic falls outside the range [+1.96, −1.96], at a significance level of 5% ($\alpha = 0.05$). A positive DM value indicates that the federated model will have less forecast loss (better performance), whereas a negative value indicates that the centralized model will have better performance.

5. Results and discussion

In order to position the SAGE-ET ConvLSTM and federated results in context, we refer to our previous studies [16,31] as the baseline, as they provide a comprehensive comparative analysis of deep learning architectures, including the convolutional LSTM, and other machine learning models for estimating ET_0. Since SAGE-ET extends this previous work into a federated learning framework, the original ConvLSTM architecture is directly reused here, where the feasibility of decentralized model learning between distinct climatic regions is demonstrated. This section showcases the findings of the experiment, cementing the need for generalization in ET_0 calculations by comparing the results of a centralized ConvLSTM model with a decentralized federated approach. The section commences by examining the cross-regional efficiency of the ConvLSTM model across four different regions with two sets of climatic features (3 and 5). Following that, the performance of the federated model is analyzed. The federated model sought to generalize the ET_0 forecast across all four regions.

5.1. Baseline performance of centralized ConvLSTM across climate zones

This subsection analyses the performance of the ConvLSTM model when model training was conducted using single-region data and tested on different regional data. We analyze this in two parts: performance evaluation with three climatic features (T_{max}, T_{min}, and R_s), bare minimum features needed in data-scarce regions for an effective ET_0 forecasting, and then with five climatic features (T_{max}, T_{min}, R_s, U_2, and R_h), for a more comprehensive study.

5.1.1. Impact of three climatic features on model performance

Initially, the ConvLSTM model was trained on three climatic features (T_{max}, T_{min}, and R_s). The statistical metrics like R^2, RMSE, and MAE were computed each time, taking one region at a time for training and then testing it on every region. Table 3 shows that the model performs remarkably well when tested on the data from the same region it was trained. It also performed decently on regions that were similar in their climatic conditions to those of the training region, but performed poorly when tested on regions with different climates. For instance, when the model was trained on data from the semi-arid region, the model achieved a high R^2 score of 0.89 as well as lower MAE and RMSE scores of 0.05 $mm\,d^{-1}$ and 0.06 $mm\,d^{-1}$ respectively, when tested on the semi-arid region and R^2 score of 0.76, MAE and RMSE scores of 0.07 $mm\,d^{-1}$ and 0.09 $mm\,d^{-1}$ respectively for arid region, but the error scores increased and the R^2 score dropped when the test region was changed to humid sub-tropical (R^2 score of 0.66, MAE score of 0.09 $mm\,d^{-1}$ and RMSE score of 0.11 $mm\,d^{-1}$) or tropical wet and dry (R^2 score of 0.73, MAE score of 0.07 $mm\,d^{-1}$ and RMSE score of 0.09 $mm\,d^{-1}$), which have different climates than the test region. This reflects the difficulty of applying a region-specific model to forecast ET_0 in regions with different climatic conditions.

5.1.2. Impact of five climatic features on model performance

In order to improve the model's performance, five climatic features were used, namely T_{max}, T_{min}, R_s, U_2, and R_h. Table 4 shows the model performance. A clear improvement in metrics is observed across all the regions, thus demonstrating that when comprehensive data is provided, model performance increases.When Tables 3 and 4 are compared, we

Table 3
Performance metrics of the centralized ConvLSTM model with three climatic features for different training regions.

Test region	MAE (mm/day)	RMSE (mm/day)	R^2
(a) Train region: Arid			
Arid	0.0635	0.0830	0.8343
Humid subtropical	0.0959	0.1188	0.6198
Semi-Arid	0.0584	0.0778	0.8554
Tropical wet and dry	0.0595	0.0818	0.8236
(b) Train region: humid subtropical			
Arid	0.1390	0.1688	0.3168
Humid subtropical	0.0241	0.0333	0.9701
Semi-arid	0.1196	0.1397	0.5347
Tropical wet and dry	0.0964	0.1229	0.6023
(c) train region: semi-arid			
Arid	0.0766	0.0986	0.7662
Humid subtropical	0.0952	0.1124	0.6601
Semi-arid	0.0503	0.0660	0.8958
Tropical wet and dry	0.0751	0.0997	0.7387
(d) Train Region: Tropical Wet and Dry			
Arid	0.0953	0.1202	0.6528
Humid subtropical	0.0677	0.0901	0.7814
Semi-arid	0.0900	0.1118	0.7017
Tropical wet and dry	0.0372	0.0517	0.9296

Table 4
Performance metrics of the centralized ConvLSTM model with five climatic features for different training regions.

Test Region	MAE (mm/day)	RMSE (mm/day)	R^2
(a) Train region: arid			
Arid	0.00738	0.00964	0.9977
Humid subtropical	0.1011	0.1291	0.5516
Semi-Arid	0.0239	0.0294	0.9793
Tropical wet and dry	0.0820	0.0875	0.7984
(b) Train region: humid subtropical			
Arid	0.1201	0.1432	0.5076
Humid subtropical	0.0134	0.0176	0.9916
Semi-Arid	0.1011	0.1185	0.6648
Tropical wet and dry	0.0574	0.0754	0.8501
(c) Train region: semi-arid			
Arid	0.0234	0.0294	0.9791
Humid subtropical	0.0870	0.1095	0.6773
Semi-Arid	0.0081	0.0104	0.9974
Tropical wet and dry	0.0609	0.0651	0.8886
(d) Train region: tropical wet and dry			
Arid	0.0933	0.1007	0.7562
Humid subtropical	0.0794	0.0988	0.7371
Semi-Arid	0.0859	0.0962	0.7792
Tropical wet and dry	0.0055	0.0072	0.9986

see a clear reduction in MAE scores. This drop is not due to any scaling inconsistency but rather due to the input feature setup. The model whose results are described under Table 3 uses only three climatic features (T_{max}, T_{min}, R_s), representing a minimalistic setup that can be used for regions with data scarcity, whereas Table 4 uses the optimal set of five features (T_{max}, T_{min}, R_s, U_2, R_h) that comprehensively reflect the governing variables of the FAO-56 Penman–Monteith formulation. In the earlier study [16], multiple architectures and feature set combinations were evaluated, concluding that the five-feature set (T_{max}, T_{min}, R_s, U_2, R_h) provides the most precise and consistent ET_0 predictions, while the three-feature set (T_{max}, T_{min}, R_s) yields the best possible performance in data-scarce conditions, with ConvLSTM showing superior accuracy. Therefore, the reduction in MAE between Tables 3 and 4 is physically consistent and reflects the contribution of the additional features to improving prediction accuracy.

However, the performance gap during cross-regional testing is still a problem. For instance, when training with data from the semi-arid region, the model's R^2 score for the semi-arid testing region increased from 0.89 to 0.99 and from 0.76 to 0.97 for the arid region, but for regions with a contrasting climate, such as the sub-humid region or tropical wet and dry region, it remained low at 0.67 and 0.88, respectively. Table 4 (a–d) depicts that the model trained and tested on the same climatic region—such as the Tropical Wet and Dry model when tested on Tropical Wet and Dry data—shows excellent performance (MAE = 0.0055 mm day^{-1}, $RMSE$ = 0.0072 mm day^{-1}, R^2 = 0.9986). A similar pattern is observed in the Arid, Semi-Arid, and Humid Subtropical models, where the same-region training and testing yield nearly perfect R^2 scores, with the second-best results obtained for regions that are climatologically similar to the training region (e.g., a model trained on Semi-Arid data performs best on Semi-Arid testing data and second-best on Arid data). This pattern confirms that the high same-region accuracy is not an anomaly arising from data leakage or overfitting but rather a systematic outcome. To avoid information leakage, the training, validation, and test splits were defined in strict chronological order and kept fully independent. The strong correlation between the climatic variables (T_{max}, T_{min}, R_s, U_2, R_h) and ET_0 in regions with homogeneous climatic conditions allows the ConvLSTM to accurately emulate the FAO-56 Penman–Monteith process. These results are consistent with the findings of our earlier work [16], where ConvLSTM models using the same three- and five-feature configurations achieved the most precise and physically consistent ET_0 estimates among multiple architectures and feature sets.

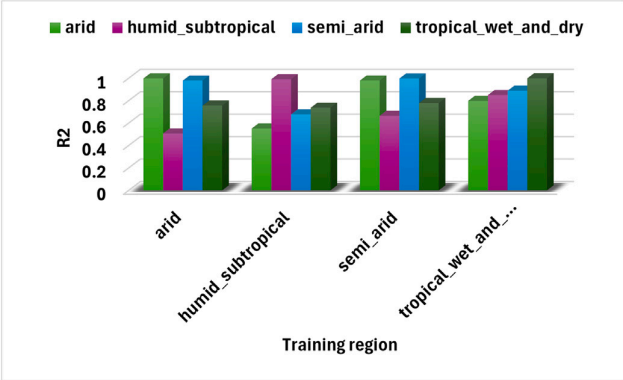


Fig. 3. Cross-regional effectiveness of the centralized ConvLSTM model in terms of R^2 for different training and testing regions.

Fig. 3 shows the R^2 for every testing and training combination of regions. It can be clearly observed that there is a dire need for a model that generalizes forecasting of ET_0 and is able to handle the climate diversity of different regions, which led to the model proposed in this paper.

5.2. Assessment of SAGE-ET in diverse climate zones

This subsection will dive into a detailed evaluation of the SAGE-ET framework, which was also trained with three climatic features (T_{max}, T_{min}, and R_s) and five climatic features (T_{max}, T_{min}, R_s, U_2, and R_h).

The federated model handled data from all four regions together, thus improving generalization, which resulted in better model performance across all the regions. The present work is primarily intended to serve as a feasibility study that substantiates the validity of SAGE-ET for ET_0 generalization. Here, the SAGE-ET's scalability is illustrated in concept with the study centering on four distinct climatic regions, each with two data collection stations to record regional diversity in a regulated and explainable approach. This setup is enough to confirm the viability and robustness of the framework under varying conditions. The architecture is also by default capable of adding and handling more clients

Table 5
Comparison of overall metrics for three and five climatic features.

Features	MAE (mm/day)	RMSE (mm/day)	R^2
5 Climatic features	0.0403	0.0527	0.9295
3 Climatic features	0.0651	0.0841	0.8210

Table 6
Performance metrics of Sage-ET across regions using three climatic features.

Region	MAE (mm/day)	RMSE (mm/day)	R^2
Arid	0.0807	0.1010	0.7532
Humid subtropical	0.0663	0.0828	0.8114
Semi-Arid	0.0658	0.0807	0.8450
Tropical wet and dry	0.0474	0.0684	0.8760

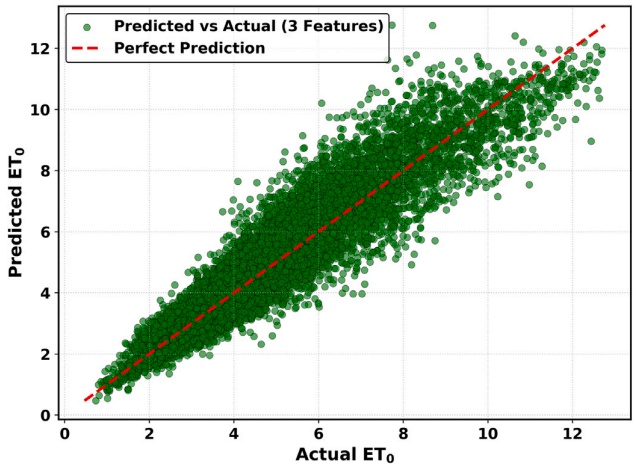


Fig. 4. Scatter plot depicting SAGE-ET predictions vs. Actual ET_0 using three climatic variables.

and stations without changing the logic or protocol of the setup, thus establishing SAGE-ET as a stepping stone for large-scale future rollout of federated learning in meteorological applications.

5.2.1. SAGE-ET performance with three climatic inputs

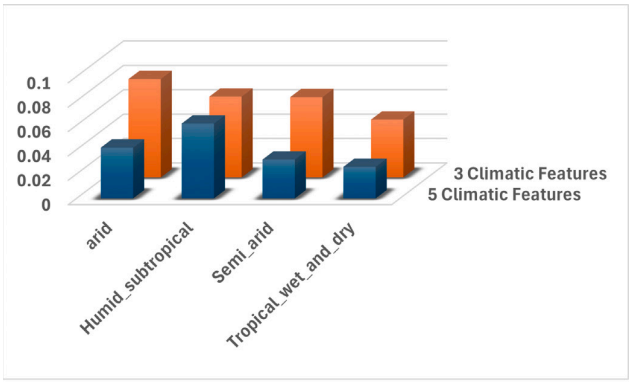
The proposed framework was first trained using three climatic features, T_{max}, T_{min}, and R_s to forecast ET_0 across the four regions. Table 5 shows how the model performs overall with an R^2 score of 0.82, 0.06 $mm\,d^{-1}$ MAE, and 0.08 $mm\,d^{-1}$ RMSE score. The model shows a very decent fit, with the ability to perform up to the mark in all the regions, thus supporting the generalization as shown in Table 6. The scatter plot (Fig. 4) demonstrates the close positive correlation between the predicted and observed values, cementing the model's ability to gauge the general trend within each region.

Despite this, there is a lot of variation in model performance in various regions due to each region's vastly distinct climate. The use of only three features can limit the capacity of the proposed SAGE-ET framework to capture the heterogeneous climatic patterns across the region. Therefore, incorporating a more comprehensive feature set can better represent the local complexities of diverse regions and also improve model generalization and predictive accuracy in such environments. The next subsection explores SAGE-ET performance using a five-feature set of climate data.

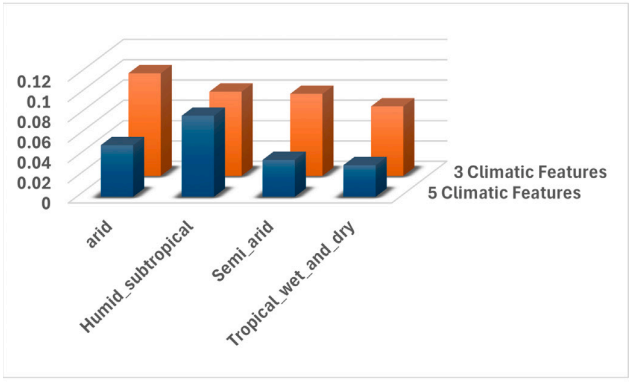
5.2.2. SAGE-ET performance with five climatic inputs

In this section, SAGE-ET framework was trained using a feature set of five climatic parameters—T_{max}, T_{min}, R_S, U_2, and R_h. The inclusion of these additional features helps the model to gain deep insight into the various climate factors that influence ET_0, thus improving the accuracy and responsiveness of the estimation to regional climate variation. Fig. 5 highlights this improvement, showing noticeable increases in the R^2 scores, MAE, and RMSE scores of the model with a set of 5 features compared to the model with a set of 3 features. The suggested SAGE–ET of five factors had a total R^2 score with a value of 0.929 (Table 5), which demonstrates improved predictions in the four regions.

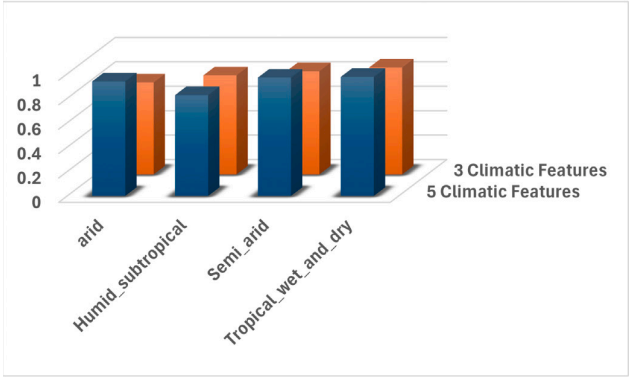
The enhanced generalization is further established by the cross-regional predictions of SAGE-ET, having a high level of R^2 and lowering scores of MAE and RMSE across the board (Table 7). Nonetheless, it is possible to notice that the humid subtropical region is an outlier. The effective performance of the organization is despite the model continuing to have a low level of accuracy as compared to other regions. This



(a) MAE for 3 and 5 climatic features across regions.



(b) RMSE for 3 and 5 climatic features across regions.



(c) R^2 for 3 and 5 climatic features across regions.

Fig. 5. Performance metrics (MAE, RMSE, and R^2) of the proposed SAGE-ET model across various climatic regions.

Table 7
Performance metrics of SAGE-ET across regions using five climatic feature.

Region	MAE (mm/day)	RMSE (mm/day)	R^2
Arid	0.0416	0.0508	0.9377
Humid subtropical	0.0613	0.0791	0.8255
Semi-Arid	0.0318	0.0363	0.9687
Tropical wet and dry	0.0258	0.0310	0.9744

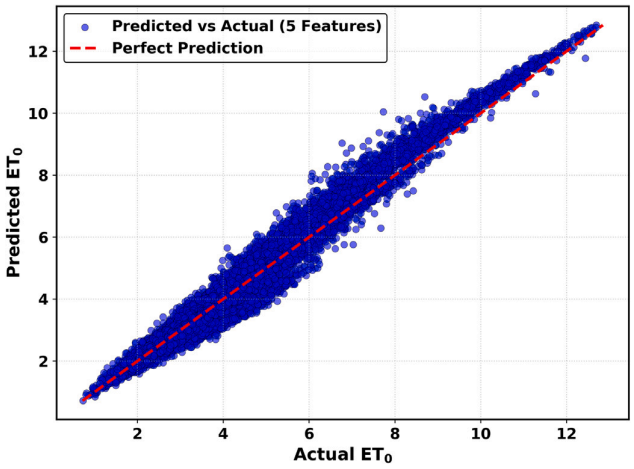
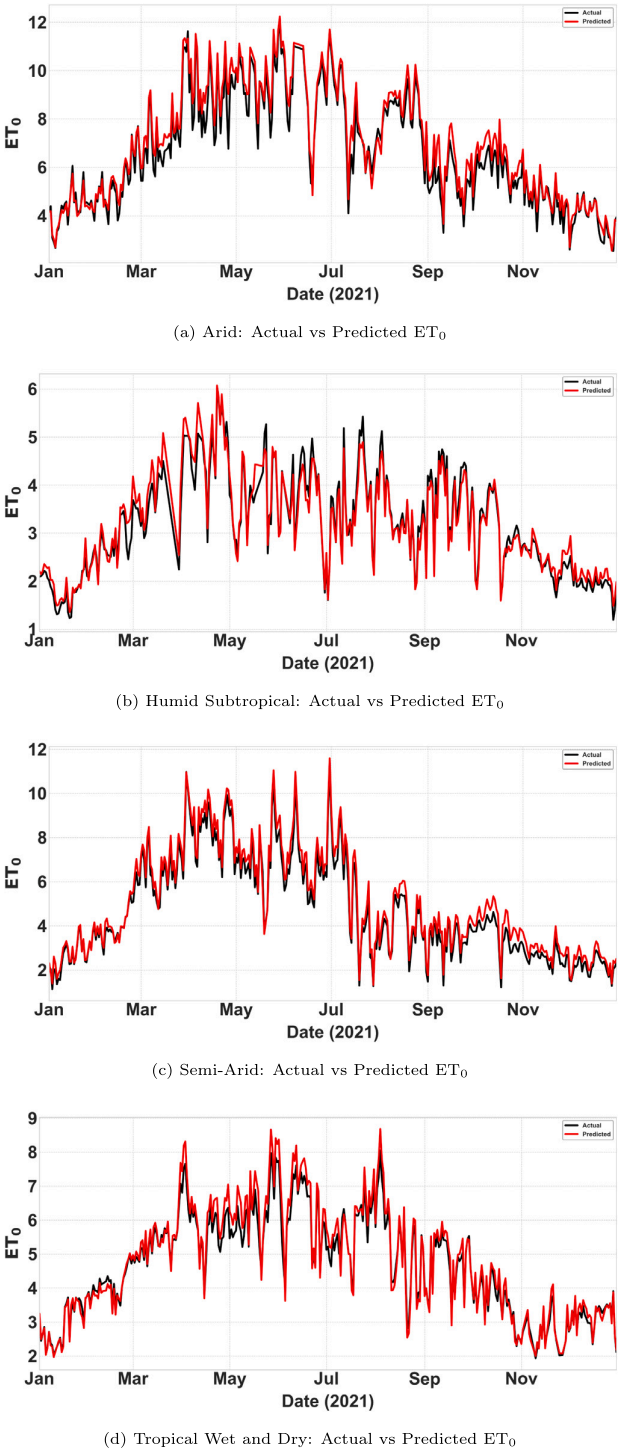


Fig. 6. Scatter plot depicting SAGE-ET predictions vs. Actual ET_0 using five climatic variables.

is because of the continuous humidity and cloud cover mainly caused by the weather conditions, making solar radiation hard to record [32]. Solar radiation (R_s) has one of the strongest influences on the estimation of ET_0 (as indicated by the heat map in Fig. 2); any discrepancies in its measurements compromise the calculations of ET_0, which would lead to decreased forecasting accuracy in that area. On the other hand, the Tropical Wet and Dry region exhibits a more predictable seasonal cycle, enabling the model to better capture the patterns that drive ET_0 prediction with higher accuracy. Furthermore, the dataset for the Humid Subtropical region contains a greater proportion of missing and noisy records—primarily due to cloud cover interfering with the measurement of solar radiation—which adversely affects forecasting accuracy in this region despite the preprocessing. Collectively, these factors contribute to the larger prediction errors observed in the Humid Subtropical region compared to the Tropical Wet and Dry region, even though both share broadly similar humidity levels.

However, the cumulative performance of the SAGE-ET shows a very good improvement across the region. The scatter plot (Fig. 6) shows a much tighter correlation between the predicted and observed values with much better alignment along the best-fit line, supporting the improved prediction accuracy of the model. Fig. 7 presents the line graphs of the predicted vs. observed ET_0 values for each region. The figure shows a near-perfect replication of the actual ET_0 values by the model across all regions, thus cementing the Federated model's ability to overcome the contrasts in model performance across all regions.

Even though SAGE-ET demonstrated strong cross-regional generalization overall, it does not necessarily surpass the best possible centralized model in every instance. For example, as observed in Table 4(c), the centralized model trained on the Semi-Arid dataset achieved a very high R^2 value of 0.9791 when tested on the Arid region dataset. However, when observing the SAGE-ET results for the same test region (Table 7), the model attained a slightly lower R^2 score of 0.9377



(a) Arid: Actual vs Predicted ET_0

(b) Humid Subtropical: Actual vs Predicted ET_0

(c) Semi-Arid: Actual vs Predicted ET_0

(d) Tropical Wet and Dry: Actual vs Predicted ET_0

Fig. 7. Comparison of actual and predicted ET_0 values across four climatic zones using the ConvLSTM model.

compared to the centralized model. This difference reflects the trade-off that exists between specialization locally and global flexibility. A centralized model based on the training and testing on the highly similar climatic regions, specialization can be effectively realized, and the performance can be close to optimal on a specific pair, whereas

Table 8
Performance comparison of SAGE-ET with prior hybrid DL models for ET_0 prediction..

Model type	Key examples	RMSE (mm/day)	R^2	Limitations addressed by SAGE-ET
CNN-LSTM Hybrid	CNN-LSTM [25]	0.37 – 0.61	0.89 – 0.91	Centralized overfitting; SAGE-ET uses FL for non-IID data
TCN hybrids	TCN-BiLSTM [26]	–	0.90–0.95	No privacy/scalability; SAGE-ET decentralizes training
SAGE-ET (Proposed)	**ConvLSTM-FedProx**	**0.031–0.051**	**0.94–0.97**	**N/A (superior in hetero. climates)**

SAGE-ET is supposed to ensure the same level of performance in several heterogeneous regions. In SAGE-ET, the regularizer that is used is FedProx, which absorbs modeling aggregation of solutions to a consensus (in some cases to the unanimous) cost of establishing the optimum accuracy in one area. This deliberate design choice will allow SAGE-ET to attain strong generalization in heterogeneous climatic conditions.

As in this section, it has been emphasized that the ET_0 generalization capability is better when compared to the baseline centralized ConvLSTM model. The proposed SAGE–ET framework provides a robust, privacy-conscious, and scalable ET_0 estimation solution by combining ConvLSTM networks with FedProx optimization in a federated learning scenario. To further validate the results, the comparison of the proposed SAGE-ET with past existing studies that have used hybrid deep learning models is shown in Table 8. Prior hybrid deep learning (DL) models have advanced ET_0 estimation by fusing convolutional and recurrent layers to capture spatiotemporal climate dependencies. For instance, CNN-LSTM hybrids achieved RMSE of 0.37–0.61 mm/day and R^2 of 0.89–0.91 on arid/semi-arid data, outperforming physics-based FAO-56 in data-scarce scenarios [25]. Other hybrids like TCN-BiLSTM reported R^2 of 0.90–0.95 across multi-step forecasting, yet lacked privacy in distributed agriculture data [26]. SAGE-ET advances these by integrating ConvLSTM within FedProx-FL, yielding RMSE 0.031–0.051 mm/day and R^2 0.94–0.97 across four non-IID climates (arid to humid subtropical), surpassing centralized hybrids' generalization. Unlike prior works assuming IID data, FedProx mitigates client drift, enabling scalable, privacy-aware ET_0 in heterogeneous regions—a gap unaddressed in hybrids [27,28].

The limitations in capturing cross-regional patterns of the single-region ConvLSTM model were addressed through the use of Federated Learning framework. By leveraging data from multiple regions, the proposed approach enabled the development of a more accurate and generalizable model for predicting ET_0 across diverse regions.

As a benchmark evaluation, the impact of choosing federated optimization strategy is further assessed by comparing Fed-Avg and FedProx under identical environment settings. Table 9 compares the region-wise performance of FedAvg and FedProx, the results clearly indicate that FedProx demonstrates better performance than FedAvg across all climate regions, which is reflected by higher R^2 and lower RMSE and MAE. Ablation (Table 9) confirms FedProx's 15% generalization gains over FedAvg baselines. The gap in performance is more prominent in arid and semi-arid regions; in these regions, statistical and climatic heterogeneity is more severe. While FedAvg is widely adopted as a baseline for federated learning, its performance diminishes under non-IID data distributions, which leads to increased prediction errors. In contrast, FedProx demonstrates better convergence, improved generalized learning behaviour, and showcases resistance to client drift.

Table 10
Diebold–Mariano (DM) test results comparing the predictive accuracy of the federated and centralized models across climatic regions. Positive DM values indicate that the federated model achieved lower forecast losses.

Region	DM (Mean ± SD)	Mean p–value
Arid	6.86 ± 1.07	0.0746
Semi–Arid	6.98 ± 1.00	0.0109
Humid subtropical	15.96 ± 1.52	0.1232
Tropical wet and dry	10.46 ± 1.11	<0.001

Federated learning is being embraced in recent studies in meteorology prediction activities like rainfall forecasts and air-quality forecasting; most of these works utilize standard CNN or LSTM models and can only predict variables one at a time. In contrast, the current research has contributed to the use of a federated learning system to estimate ET_0, a multivariate, physically-grounded process. In the proposed SAGE-ET_0, the ConvLSTM, which is effective at learning spatiotemporal climatic dependencies, is coupled with FedProx, a federated optimization approach implemented to reduce client drift corresponding to heterogeneous data. SAGE-ET also focuses on cross-regional generalization across four major climatic regions of India, while maintaining consistency with the FAO-56 Penman–Monteith formulation. This work highlights the benefits of federated learning structures to be exploited as de facto, privacy-aware, and scalable environmental modeling frameworks.

5.3. Statistical comparison between centralized and federated models

Diebold–Mariano (DM) test was conducted in order to statistically compare the differences in predictive accuracy between federated SAGE–ET and centralized ConvLSTM models. The results are given in Table 10 and indicate that there are positive DM statistics across all climatic regions, which means that the federated model attained less forecast losses than in the centralized model.

The DM values fell within 6.86 to 15.96 which is more than the critical interval of ±1.96 at a 5% significance level. This proves that the null hypothesis is false, and so the null hypothesis which says both the model have the same predictive power may be rejected everywhere as both models shows significant changes in performance. The highest DM values had been found in the tropical wet and dry climates and the humid-subtropical region (10.46 and 15.96, respectively), which supports the federated model of generalizability in different and extreme climatic conditions to a great extent.

Table 9
Region-wise performance comparison of FedAvg and FedProx.

Region	R^2 (Fed- Avg)	R^2 (FedProx)	RMSE (Fed- Avg)	RMSE (FedProx)	MAE (Fed- Avg)	MAE (FedProx)
Tropical wet & Dry	0.9682	0.9744	0.0346	0.0310	0.0299	0.0258
Semi-Arid	0.9488	0.9687	0.0464	0.0363	0.0432	0.0318
Arid	0.9101	0.9377	0.0610	0.0508	0.0525	0.0416
Humid subtropical	0.8627	0.9255	0.0847	0.0791	0.0647	0.0615

There was also statistically significant improvement in semi-arid region (DM = 6.98, $p = 0.0109$), which shows that the federated model generated significantly smaller forecast errors than the centralized ConvLSTM. Although the arid climate tended to have a comparatively high p-value ($p = 0.0746$), the positive DM statistic (6.86) indicates that the predictive skill is gradually improving.

The trend in using p-values in the regions also indicates the legitimacy of the findings. Areas where $p < 0.05$, like the semi-arid and tropical wet and dry climates, furnish ample statistical proof of the superiority of the federated model. The arid regions and the humid subtropical do not fall under the 5% cutoff, but still show positive DM values, meaning that performance improvements are systematic without necessarily being significant formally. Overall, these results indicate that the federated SAGE–ET model is statistically verifiable as more effective than the centralized ConvLSTM model, supporting its enhanced capacity to generalize under non-IID climatic conditions.

In general, the results confirm that the federated SAGE–ET framework exhibits statistically significant and regionally consistent improvements in reference evapotranspiration estimation compared to the centralized ConvLSTM model. The prevalence of low p–values and consistently positive DM statistics provides strong evidence that the superior performance of the federated model is not incidental. These findings further reinforce the effectiveness of federated learning in generalizing under non-IID and heterogeneous climatic conditions.

6. Conclusion and future work

This study proposed **SAGE-ET**, a novel federated learning framework that employs **FedET_0-Prox**, a novel federated learning model that combines FedProx optimization and a ConvLSTM architecture to generalize reference evapotranspiration (ET_0) to enable confidentiality-preserving, robust and accurate prediction of ET_0. The results illustrate that SAGE-ET performs better than a centralized ConvLSTM framework in arid, semi-arid, tropical wet and dry, and humid subtropical regions, exhibiting higher R^2 scores and lower MAE and RMSE values. This improved performance effectively addresses the shortcomings of traditional, localized machine learning models and illustrates the model's resilience in the face of diverse input and regional variation. SAGE-ET offers a promising direction for secure, scalable, adaptable, and privacy-aware generalized ETo modeling.

This work builds the foundation for future research on decentralized climate modeling and cross-regional forecasting. This study was primarily intended to serve as a feasibility study that substantiates the validity of SAGE-ET for ET_0 generalization. Here, federated learning was applied on four distinct regions (two data stations each). Future work will explore the scaling of the federated architecture to a bigger group of clients, with more climate-related diversity, greater data magnitude, and hardware capabilities. This will help further enhance SAGE-ET's capabilities as an adaptable, climate data-driven learning architecture by enabling detailed evaluation of communicational latency, robustness of parameter fusion, and how convergence happens under practical scenarios. SAGE-ET ensures decentralization of raw data, but its privacy can be further enhanced by incorporating secure aggregation schemes, wherein the central server receives only aggregated model updates rather than individual client parameters, effectively eliminating the possibility of client data leakage. Furthermore, differential privacy techniques can be employed to inject controlled statistical noise into local updates before transmission, thereby providing strict guarantees against information leakage. With this feasibility study, the research also forms a basis for further work in decentralized climate modeling and interregional forecasting. In addition to being methodologically validated, the proposed SAGE-ET framework illustrates the feasibility of applying federated learning to climate-aware agriculture. It highlights the potential for creating privacy-conscious and locally adaptive irrigation decision-making systems, demonstrating that ET_0 can be collaboratively estimated across different climatic regions without centralized data exchange. Although large-scale implementation is beyond the scope of the current research, it is important to emphasize that the demonstrated feasibility of SAGE-ET can be envisioned as part of local irrigation scheduling, water-allocation planning, and climate-resilient agricultural programs. This approach can support sustainability initiatives such as the National Water Mission in India and the FAO's sustainable water-use guidelines by enhancing data-driven water management processes and improving agricultural resilience to climatic variability.

Despite of the promising results, there are possibilities for improvement in the existing proposed framework. A sound technique is to include secure aggregation techniques and differential privacy in the proposed framework. These safeguards help enhance the privacy assurances of federated learning in scenarios where updates to the model could potentially leak sensitive data. In order to overcome the difficulties of client heterogeneity and the limited computing/network resources, future research should also explore communication-efficient federated optimization algorithms and adaptive client participation strategies, in addition to integrating multimodal data, such as crop stages, soil moisture, satellite-based ET_0 estimates, and a crucial future goal. Lastly, introducing and verifying SAGE–ET in the real world. Agricultural systems, by cross-disciplinary collaboration and real-time feedback, can pave the way for wider use in operational water management practices.

7. Limitations

SAGE-ET demonstrated good generalization ability for most of the regions of the weather zone, but relatively low generalization ability for humid subtropical regions. This is the result of the increased variability of climatic parameters, such as rainfall, cloud cover, or humidity, and possible data quality limitations in these stations. In addition, implementation of the federated frameworks in the real world is threatened by resource distribution, creating security vulnerabilities, such as communication delays, limited bandwidth, and asymmetric resource distribution. Potential improvement directions include using better personalization of the global model, aggregation methods that include heterogeneity on the client side, and an efficient communication framework in low-bandwidth settings.

CRediT authorship contribution statement

Gitika Sharma: Writing – review & editing, Supervision, Project administration. **Aayush Tripathi:** Writing – original draft, Software, Methodology. **Atharva I.P. Agrawal:** Writing – original draft, Software. **Himanshu Sharma:** Writing – original draft, Software, Methodology. **Laxmi Kant Sharma:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank the India Meteorological Department (IMD), Pune, for providing the climate dataset used in this study.

References

[1] Food and Agriculture Organization, Crop evapotranspiration—guidelines for computing crop water requirements, available at: https://www.fao.org/4/x0490e/x0490e06.htm#TopOfPage, 2025.
[2] A. Ayaz, M. Rajesh, S.K. Singh, S. Rehana, Estimation of reference evapotranspiration using machine learning models with limited data, AIMS Geosciences 7 (3) (2021) 268–290.
[3] L. Huang, Y. Luo, J.M. Chen, Q. Tang, T. Steenhuis, W. Cheng, W. Shi, Satellite-based near-real-time global daily terrestrial evapotranspiration estimates, Earth Syst. Sci. Data 16 (2024) 3993–4019, https://doi.org/10.5194/essd-16-3993-2024
[4] A. Raza, Y. Hu, M. Shoaib, M.K. Abd Elnabi, M. Zubair, M. Nauman, et al., A systematic review on estimation of reference evapotranspiration under Prisma guidelines, Pol. J. Environ. Stud. 30 (2021) 5413–5422.

[5] M. Kumar, Y. Agrawal, S. Adamala, Pushpanjali, A.V.M. Subbarao, V.K. Singh, et al., Generalization ability of bagging and boosting type deep learning models in evapotranspiration estimation, Water 16 (16) (2024) 2233.

[6] I. Ghiat, H.R. Mackey, T. Al-Ansari, A review of evapotranspiration measurement models, techniques and methods for open and closed agricultural field applications, Water 13 (18) (2021) 2523, https://doi.org/10.3390/w13182523

[7] H. Shi, X. Cai, Extrapolability improvement of machine learning-based evapotranspiration models via domain-adversarial neural networks, Environ. Model. Softw. 187 (2025) 106383.

[8] Q. Li, C. Zhang, Z. Wei, X. Jin, W. Shangguan, H. Yuan, et al., Advancing symbolic regression for earth science with a focus on evapotranspiration modeling, npj Clim. Atmos. Sci. 7 (1) (2024) 321.

[9] K.R. Žalik, M. Žalik, A review of federated learning in Agriculture, Sensors 23 (23) (2023) 9566.

[10] U. Akbulut, M.A. Cifci, Z. Aslan, Hybrid modeling for stream flow estimation: integrating machine learning and federated learning, Appl. Sci. 13 (18) (2023) 10203.

[11] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H.B. McMahan, S. Patel, D. Ramage, A. Segal, K. Seth, Practical secure aggregation for privacy-preserving machine learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), 2017, pp. 1175–1191, https://doi.org/10.1145/3133956.3133982

[12] T. Li, A.K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, in: Proceedings of Machine Learning and Systems, vol. 2, 2020, pp. 429–450.

[13] M.J.T. Kathen, P. Johnson, I.J. Flores, Aquafel-pso: A monitoring system for water resources using autonomous surface vehicles based on multimodal pso and federated learning, arXiv preprint arXiv:2211.15217, 2022.

[14] A. Mukherjee, R. Buyya, Federated learning architectures: a performance evaluation with crop yield prediction application, Softw. Pract. Exp. 55 (7) (2025) 1165–1184.

[15] A. Alwabli, Federated Learning for privacy-preserving AIR quality forecasting using IOT sensors, Eng. Technol. Appl. Sci. Res. 14 (4) (2024) 16069–16076, https://doi.org/10.48084/etasr.7820

[16] G. Sharma, A. Singh, S. Jain, A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data, Neural Comput. Appl. 34 (2022) 1–20.

[17] H.B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, Proc. Mach. Learn. Res. 54 (2017) 1273–1282.

[18] T. Li, A.K. Sahu, A. Talwalkar, V. Smith, Federated learning: challenges, methods, and future directions, IEEE Signal Process. Mag. 38 (3) (2021) 50–60.

[19] P. Kairouz, H.B. McMahan, B. Avent, A. Bellet, M. Bennis, A.N. Bhagoji, et al., Advances and open problems in federated learning, Found. Trends Mach. Learn. 14 (1–2) (2021) 1–210.

[20] S.P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, A.T. Suresh, Scaffold: stochastic controlled averaging for federated learning, Proc. Mach. Learn. Res. 139 (2021) 5132–5143.

[21] T. Li, A.K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Fedprox: federated optimization in heterogeneous networks, in: Proceedings of Machine Learning and Systems, 2020, pp. 429–450.

[22] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, Y. Khazaeni, Tackling the objective inconsistency problem in heterogeneous federated optimization, in: Advances in Neural Information Processing Systems, 2021, pp. 7611–7623.

[23] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, et al., Practical secure aggregation for federated learning on user-held data, in: Proceedings of the IEEE Symposium on Security and Privacy, 2021, pp. 1352–1369.

[24] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, et al., A hybrid approach to privacy-preserving federated learning, in: Proceedings of the ACM Conference on Computer and Communications Security, 2021, pp. 1556–1569.

[25] M. Tang, Z. Ma, P. Ai, T. Heng, Y. Ma, Multi-step-ahead forecasting of daily reference evapotranspiration using hybrid deep learning models for the Taklamakan Desert Oasis, J. Hydrol.: Reg. Stud. 61 (2025) 102663.

[26] Z. Gao, X. Zhang, Z. Gao, B. Zhou, Y. Liu, D. Liu, et al., An interpretable hybrid TCN-BiLSTM model for reference evapotranspiration prediction, Water Resour. Manag. 39 (2025) 1–23.

[27] T. Li, A.K. Sahu, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, Proc. Mach. Learn. Syst. 2, fedProx original paper[file:1] (2020).

[28] Anonymous, Fedbchain: A blockchain-enabled federated learning framework for improving deepconvlstm, [web:8] arXiv:2407.21282, 2024.

[29] Flower Documentation, Federated learning baselines: FedProx, Available at: https://flower.ai/docs/baselines/fedprox.html (accessed 18 Feb 2026).

[30] X.-T. Yuan, P. Li, On convergence of fedprox: local dissimilarity invariant bounds, non-smoothness and beyond, Adv. Neural Inf. Process. Syst. 35, arXiv:2206.05187 (2022) 22639–22652, arXiv preprint https://arxiv.org/abs/2206.05187.

[31] G. Sharma, A. Singh, S. Jain, Deepevap: Deep reinforcement learning based ensemble approach for estimating reference evapotranspiration, Appl. Soft Comput. 125 (2022) 109113.

[32] B. Jamil, N. Akhtar, Comparison of empirical models to estimate monthly mean diffuse solar radiation from measured data: case study for humid-subtropical climatic region of India, Renew. Sustain. Energy Rev. 77 (2017) 1326–1342, https://doi.org/10.1016/j.rser.2017.03.027

[33] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.

[34] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, 1977.

[35] J. Chen, X. Cao, S. Wang, G. Joshi, H.V. Poor, On the importance and applicability of pre-training for federated learning, arXiv preprint arXiv:2206.11488, 2022, https://arxiv.org/abs/2206.11488.

[36] Q. Yu, X. Li, Y. Liu, Pre-training in federated Learning: a comprehensive study on communication efficiency and model robustness, IEEE Trans. Neural Netw. Learn. Syst. 33 (5) (2022) 2104–2115.

[37] H. Wang, Z. Kaplan, D. Niu, B. Li, Optimizing federated learning with pre-training for non-Iid data, J. Mach. Learn. Res. 22 (1) (2021) 1–30.

[38] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A.Y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 54, 2017, pp. 1273–1282.

[39] F. Sattler, S. Wiedemann, K.R. Müller, W. Samek, Robust and communication-efficient federated learning from non-Iid data, IEEE Trans. Neural Netw. Learn. Syst. 31 (9) (2019) 3400–3413.

[40] Y. Zhao, L. Li, E. Kruglyakova, Z. Zhao, J. Duchi, Federated learning with non-IID data, arXiv:1806.00582, 2018.

[41] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82.

[42] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econ. Stat. 13 (3) (1995) 253–263.

[43] A.J. Patton, F.X. Diebold, Discussion of comparing predictive accuracy, twenty years later, J. Bus. Econ. Stat. 32 (2014) 12–16.

[44] W.K. Newey, K.D. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55 (3) (1987) 703–708.

[45] H. Chen, Q. Wan, Y. Wang, Refined diebold–mariano test methods for the evaluation of wind power forecasting models, Energies 7 (7) (2014) 4185–4198, https://doi.org/10.3390/en7074185