Engineering Applications of Artificial Intelligence 159 (2025) 111636

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Research paper

Overcoming data scarcity: A transfer learning framework with fine-tuned neural networks and multi-sensor satellite image fusion for soil moisture estimation

Abhilash Singh [a,b,*], M. Niranjannaik [a], Kumar Gaurav [a,**]

[a] Fluvial Geomorphology and Remote Sensing Laboratory, Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, Bypass road, Bhopal, 462066, Madhya Pradesh, India

[b] School of Mathematics, Faculty of Engineering and Physical Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom

ARTICLE INFO

Dataset link: https://abhilashsingh.net/codes.html

Keywords:
Transfer learning
Deep learning
Soil moisture
Sentinel-1/2

ABSTRACT

Training deep learning (DL) models requires extensive data, particularly for soil moisture prediction, where large volumes of in situ measurements are needed to prevent overfitting. To address this challenge, we propose a customised transfer learning framework that adapts a pre-trained DL model to a new study site with a different climate. Specifically, we fine-tune a fully connected feed-forward neural network, originally trained on a large dataset from a humid subtropical region (source domain), using limited data from a semi-arid region (target domain). The proposed framework leverages nine input features extracted and generated from Sentinel-1/2 and Shuttle Radar Topographic Mission (SRTM) images through a linear data fusion technique. We systematically evaluate the performance of the proposed framework against ten benchmark algorithms. We observed that the proposed framework outperforms all benchmark algorithms, achieving a correlation coefficient (R) of 0.81, a root mean square error (RMSE) of 0.05 m^3/m^3, and a bias of 0.02 m^3/m^3 on the target domain. Particularly, this is achieved using 55% less in situ data compared to the source domain. To ensure reliability and robustness, we conduct comprehensive analyses, including error histogram, residual, uncertainty, spatial distribution, ablation, statistical, and complex time complexity analyses. Throughout each evaluation, the proposed framework consistently exhibits a reliable and robust performance. The findings of this study hold promise in facilitating accurate surface soil moisture estimation, particularly in data-scarce regions.

1. Introduction

Surface soil moisture is important to understand the dynamics between numerous processes occurring at the air-land interface (Singh and Gaurav, 2024; McColl et al., 2017). Traditional methods such as oven drying, neutron probe, gamma rays probe, time domain reflectometry, and frequency domain reflectometry are useful in estimating the point measurement of surface soil moisture (Singh et al., 2023b.). Although these techniques provide accurate measurements, they are tedious and time-consuming. Remote sensing, particularly Synthetic Aperture Radar (SAR) microwave images, is widely used for large-scale soil moisture estimation (Efremova et al., 2021; Baghdadi et al., 2004; Mirsoleimani et al., 2019; Zhu et al., 2020, 2022).

Numerous inversion-based backscattering models have been developed to estimate the surface soil moisture information from SAR images directly (Oh et al., 1992, 2002; Oh, 2004; Dubois et al., 1995; Fung et al., 1992; Attema and Ulaby, 1978). These models are directly applied and inverted using quad-polarised SAR images to estimate soil permittivity and consequently soil moisture information. A modified version of these models is employed when only dual-polarised images are available (Sahebi and Angles, 2010; Joseph et al., 2010; Singh et al., 2020; Shen et al., 2023). The major issue with the usage of these models is that they do not provide consistent performance across different climate regions. In a recent study, Singh et al. (2020, 2022) demonstrated that the modified version of these models yields satisfactory results in a semi-arid region where the soil moisture ranges between 0 and 0.30

* Corresponding author. Fluvial Geomorphology and Remote Sensing Laboratory, Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, Bypass road, Bhopal, 462066, Madhya Pradesh, India.
** Corresponding author.
E-mail addresses: sabhilash@iiserb.ac.in, a.singh4@leeds.ac.uk, abhilash.iiserb@gmail.com (A. Singh), niranjannaik@iiserb.ac.in (M. Niranjannaik), kgaurav@iiserb.ac.in (K. Gaurav).

https://doi.org/10.1016/j.engappai.2025.111636
Received 23 July 2024; Received in revised form 10 May 2025; Accepted 25 June 2025

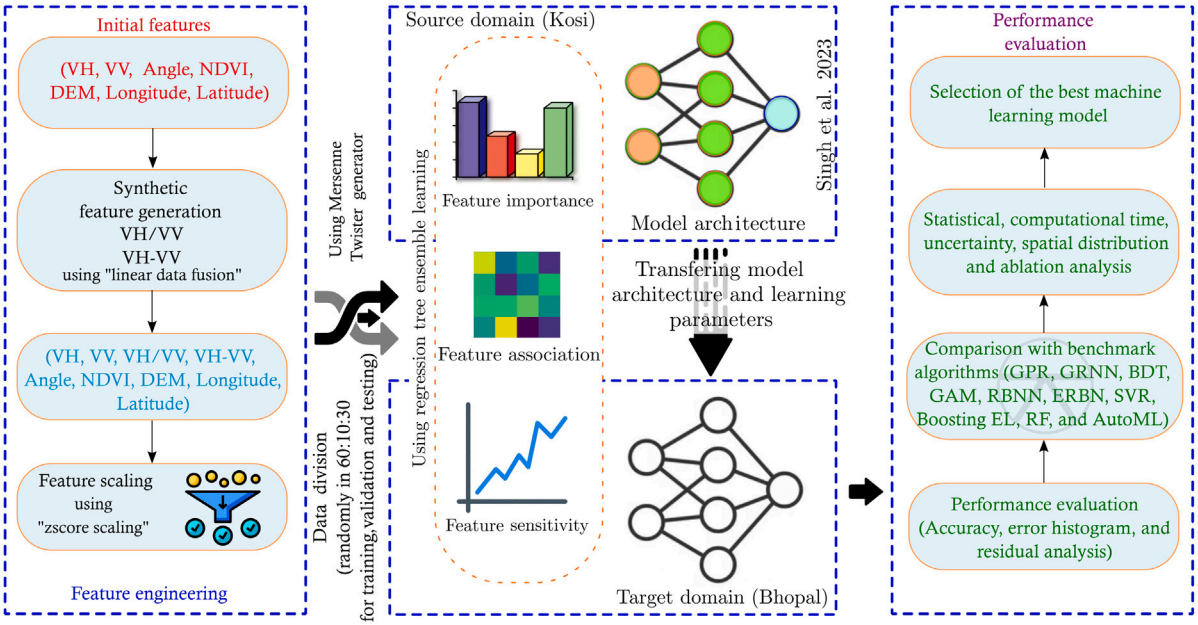

Fig. 1. Flowchart of the transfer learning framework.

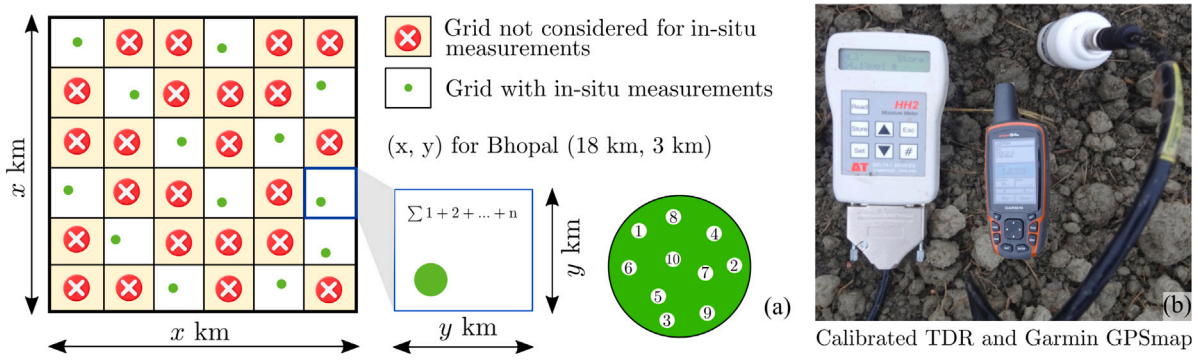

Fig. 2. (a) Schematic of the random-grid sampling design used in this study for in situ measurements. (b) A calibrated time domain reflectometer (TDR) and a handheld GPS unit were used for the measurements.

m^3/m^3, while the same model fails to estimate soil moisture accurately in humid sub-tropical regions, where the soil moisture ranges between 0.10 and 0.50 m^3/m^3. This persistent limitation is primarily rooted in a fundamental assumption of these models, which restricts their usage to regions where soil moisture is > 0.35 m^3/m^3. Subsequently, Singh and Gaurav (2023) proposed a data-driven framework to estimate soil moisture in high soil moisture regions by leveraging deep learning, data fusion, and large in situ observations. They found that the proposed framework accurately estimates surface soil moisture in humid sub-tropical regions using input features extracted from publicly available satellite images. However, testing this model across other climatic regions is very challenging due to the costs involved in collecting in situ soil moisture samples, which limits its usage in large-scale applications.

An intelligent approach to address this limitation is by transferring the learning from an existing model trained with extensive datasets, referred to as the source domain, to train a new model in a different climatic zone with limited in situ observations, known as the target domain. This transfer of parameters serves as prior knowledge for the target domain, thereby leveraging the potential of the pre-trained model to extend the application of the source domain model to larger areas, which is crucial for global soil moisture estimation. The success of transferring the learning parameters highly depends on the similarity between the tasks in the source and target domains. For instance, in the context of soil moisture estimation, the task in the source domain might involve utilising extensive in situ observations from study site A, while in the target domain, the challenge could be in estimating soil moisture with limited in situ observations from study site B. It is important to note that study sites A and B may have different climate settings. Only a few transfer learning approaches have been proposed in the field of soil moisture estimation, leaving it largely unexplored. For instance, Li et al. (2021) introduced a transfer learning framework to enhance Soil Moisture Active Passive (SMAP) soil moisture retrieval (target domain) by utilising a pre-trained model from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5)-land (source domain). In Hemmati and Sahebi (2024), a one-dimensional convolutional neural network with an integrated attention module was first trained on the global ISMN dataset and then fine-tuned on the in situ samples collected across the Karaj watershed in Iran. The input vector comprised 11 features: Sentinel-1 VV/VH backscatter statistics (mean, standard deviation, 5th and 95th percentiles), day-of-year (DOY), NDVI (from Sentinel-2), and local incidence angle. They found that the fine-tuned version of the transfer learning model outperforms several benchmark algorithms as well as the baseline transfer learning model. Recently, Zhu et al. (2024) proposed a cross-resolution transfer learning framework for the estimation of soil moisture from a single sensor (*e.g.,* Sentinel-1 images). In the source domain, the task involves the estimation of soil moisture at a coarser resolution (9 km), while in the target domain, it pertains to soil moisture estimation

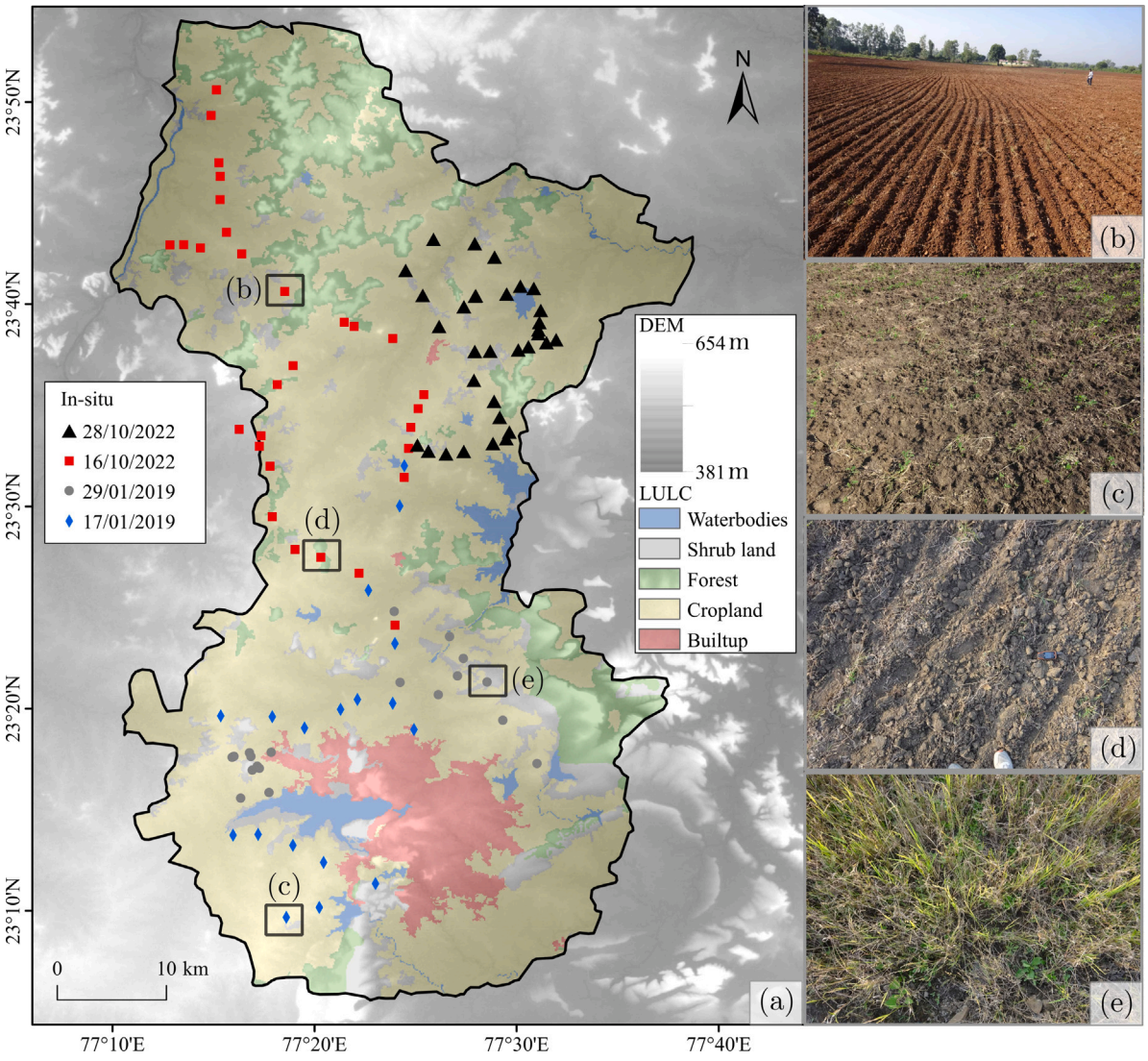

Fig. 3. Study area map of the target domain (Bhopal). Symbols of different shapes and colors represent the locations of in situ soil moisture measurements. Field photographs (b)–(e) illustrate the ground conditions at the time of measurement.

at finer resolutions (0.1, 0.5, and 1 km). In all the above-mentioned studies, the transfer learning framework was utilised to enhance the accuracy of soil moisture retrieval at the same study site. In this study, we propose a transfer learning framework to leverage the efficiency of existing pre-trained models in data-abundant regions for developing a model to estimate surface soil moisture in different climate regions, where in situ observations are limited. The climatic heterogeneity often induces shifts in feature distributions, known in the transfer-learning literature as "covariate shift" (He et al., 2024). In such scenarios, fine-tuning the baseline model is essential to account for differences in soil moisture distributions. The main contributions of this study are ordered as follows:

1. To analyse how the individual impact of each input feature changes when applied to different climate settings.
2. Assess whether model parameters or learned representations can be effectively transferred to the target domain.

This manuscript is divided into four sections. Section 2 discusses the datasets used, followed by methods used for feature importance, association, sensitivity analysis, and model development. Sections 3 and 4 present the experimental results and discuss the performance of the proposed transfer learning framework. Finally, Section 5 concludes the main findings of this study.

2. Material and methods

This section discusses the datasets used to train the proposed transfer learning framework. It covers feature preprocessing and generation, followed by the evaluation of feature importance, association, and sensitivity, and concludes with the model development process. The comprehensive methodology is depicted in Fig. 1.

2.1. In situ and satellite datasets

We conducted two field campaigns in Bhopal, one in January 2019 and the another in October 2022, to measure surface soil moisture using a calibrated time domain reflectometer (TDR). Measurements were obtained from the topsoil layer at a depth of 5 cm (Singh and Gaurav, 2023). This depth was determined based on the radar penetration depth module proposed by Singh et al. (2018, 2019), ensuring that the SAR signal effectively illuminated and recorded the backscatter value at that specific depth. More details on the calibration process can be found in Singh et al. (2020). For sampling, we employed a Universal random grid sampling approach. We superimposed a square grid with dimensions of 3 km × 3 km on the study area. Within each randomly selected grid cell, we identified a region equivalent in size to the final processed satellite pixel (60 m × 60 m) exhibiting low to

Table 1
Details of the Sentinel-1/2 images used in this study.

Sentinel-1				
Date (mm/dd/yyyy)	Pass	Polarisation	Incidence angle (Near, Far)	Pixel size (m × m)
01/17/2019	Descending	Dual (VH, VV)	(30.83°, 46.07°)	10 × 10
01/29/2019	Descending	Dual (VH, VV)	(30.83°, 46.32°)	10 × 10
10/16/2022	Descending	Dual (VH, VV)	(30.84°, 46.19°)	10 × 10
10/28/2022	Descending	Dual (VH, VV)	(30.83°, 46.18°)	10 × 10
Sentinel-2				
Date (mm/dd/yyyy)	Orbit number and direction	Bands	Wavelength (nm)	Spatial Resolution (m)
1/17/2019	62, Descending	B8, B4	646–685, 774–907	10
1/28/2022	62, Descending	B8, B4	646–685, 774–907	10

medium heterogeneity (Fig. 2). We collected in situ measurements at 99 locations and recorded their coordinates using a Garmin GPSMAP 64s handheld GPS device (Fig. 3). At these locations, we randomly collected 8–12 measurements within the pixel area and then averaged them to obtain a representative value for that location. This averaging process minimises spatial heterogeneity within the pixel, allowing for a direct point-to-pixel comparison to validate the proposed model (Ryan et al., 2017).

We used publicly accessible Sentinel-1 (SAR) and Sentinel-2 (optical) images, along with a DEM derived from the Shuttle Radar Topographic Mission (SRTM) to extract various input features for training the proposed model. We downloaded Sentinel imagery from the European Space Agency's official website and the void-filled SRTM DEM (at 1 arc-second resolution) from the U.S. Geological Survey (USGS) website (https://earthexplorer.usgs.gov). In this study, we utilised level-1 GRD dual-polarised (*i.e.,* VV+VH) Sentinel-1 and bottom-of-atmosphere corrected level-2 A Sentinel-2 images, both at a spatial resolution of 10 × 10 m. Detailed descriptions of the Sentinel images used in this study are provided in Table 1.

2.2. Image processing and feature extraction

We used the publicly available Sentinel Application Platform (SNAP v8.0) for pre-processing raw Sentinel-1 images. Our methodology comprised radiometric calibration, multi-look (with a factor of 6) correction, speckle noise reduction, and terrain correction, performed sequentially to enhance the raw Sentinel-1 images. These systematic procedures facilitated the generation of backscatter images for both VV and VH polarisations, refined to a grid size of 60 m. This grid size resulted from the multi-looking process, which averages adjacent elongated pixels into precise square pixels of 60 m by multiplying the original pixel size with the multi-looking factor. Employing the linear data fusion technique, we derived two synthetic features, VH/VV and VH-VV, from VV and VH (Ittner and Schlosser, 1996). These features exhibit higher sensitivity towards the dielectric and geometric properties of soil (Singh et al., 2021; Greifeneder et al., 2018). Additionally, the SAR incidence angle from Sentinel-1 images significantly influences satellite-derived soil moisture (Autret et al., 1989). To account for the morphology, topography, and spatial dependencies of the study area, we considered elevation and geolocation variables (*i.e.,* latitude and longitude) as input features (Lin et al., 2006; Tenenbaum et al., 2006; Murphy et al., 2009; Fathololoumi et al., 2020; Santi et al., 2016). Moreover, to address the impact of vegetation on soil moisture estimation, the Normalised Difference Vegetation Index (NDVI) from Sentinel-2 images were included (Farrar et al., 1994; Felegari et al., 2022). This involved calculating the Normalised Difference Vegetation Index (NDVI) by taking the difference between the near-infrared and red bands and dividing it by their sum, resulting in an NDVI image with a spatial resolution of 10 m. It is crucial to note that NDVI values range from −1 to +1, with higher values indicating healthier vegetation. Subsequently, to standardise all input features to a common level, we normalised them using the z-score scaling technique after applying the nearest neighbour resampling to achieve a uniform resolution of 60 m × 60 m. Since all input features are defined on a 60 m × 60 m grid, the resulting soil moisture map is likewise produced at a 60 m × 60 m resolution. Finally, we divided the entire dataset in a 60:10:30 ratio (Baier et al., 2021) for training, validation, and testing of the proposed framework using the Mersenne Twister random generator (Matsumoto and Nishimura, 1998).

2.3. Feature importance, association, and sensitivity

We assess the relevance of the input features in predicting surface soil moisture information. To accomplish this, we employ the regression ensemble tree method and gauge the importance score of each feature. Our approach involves utilising the LSBoost algorithm to train 500 regression trees (n) with a fixed learning rate (ζ) of 1 (Kumar et al., 2024). We consider traditional decision trees, specifically decision stumps, as weak learners. LSBoost operates by sequentially training one weak learner at a time while concurrently identifying its limitations to generate a new weak learner (β_i). This process is accompanied by calculating corresponding weights (η_i). Subsequently, the current model (W_i^L) is updated to address the shortcomings of the preceding weak learner (W_{i-1}^L), as shown in Eq. (1);

$$W_i^L = W_{i-1}^L + \zeta \cdot \eta_i \cdot \beta_i \qquad (i = 1, 2, 3, \ldots, n) \tag{1}$$

Iteratively, the LSBoost algorithm incorporates each weak learner into the current model, generating an ensemble of weak learners, denoted as W_m^L, ultimately constituting a single strong learner. These steps are followed by estimating the comprehensive changes in node risk caused by splitting on each feature. This estimation involves normalising the changes concerning the total number of branch nodes (N_{total}). Leveraging this normalised data, we calculate the importance score. The node changes risk (Δ_{NR}) are mathematically calculated according to Eq. (2);

$$\Delta_{NR} = \frac{P_{NR} - (C_{NR}^1 + C_{NR}^2)}{N_{total}} \tag{2}$$

where P_{NR}, C_{NR}^1, and C_{NR}^2 denote the parent and two children node risks, respectively. These values are computed by multiplying the node probability with the mean square error of the corresponding node.

Additionally, we evaluate the correlation among input features using the same regression tree ensemble learning model. The feature association matrix is computed following the method described in Singh and Gaurav (2023). Correlated features can adversely affect the stability of the machine learning model, increasing its susceptibility to uncertainty. The matrix entries represent the similarities between decision rules that split on individual observations, with values ranging from 0 to 1. Values closer to 1 indicate a high correlation between the corresponding features.

Unlike feature importance analysis, which only provides discrete importance scores without detailing the exact impact of each feature ($\boldsymbol{F} = f_1, f_2, \ldots, f_m$, where m is the total number of input features), we leveraged the mapping function (χ) of the regression tree ensemble model to quantify the impact (whether decreasing, increasing, or undulating). This was achieved by marginalising, or averaging out, the

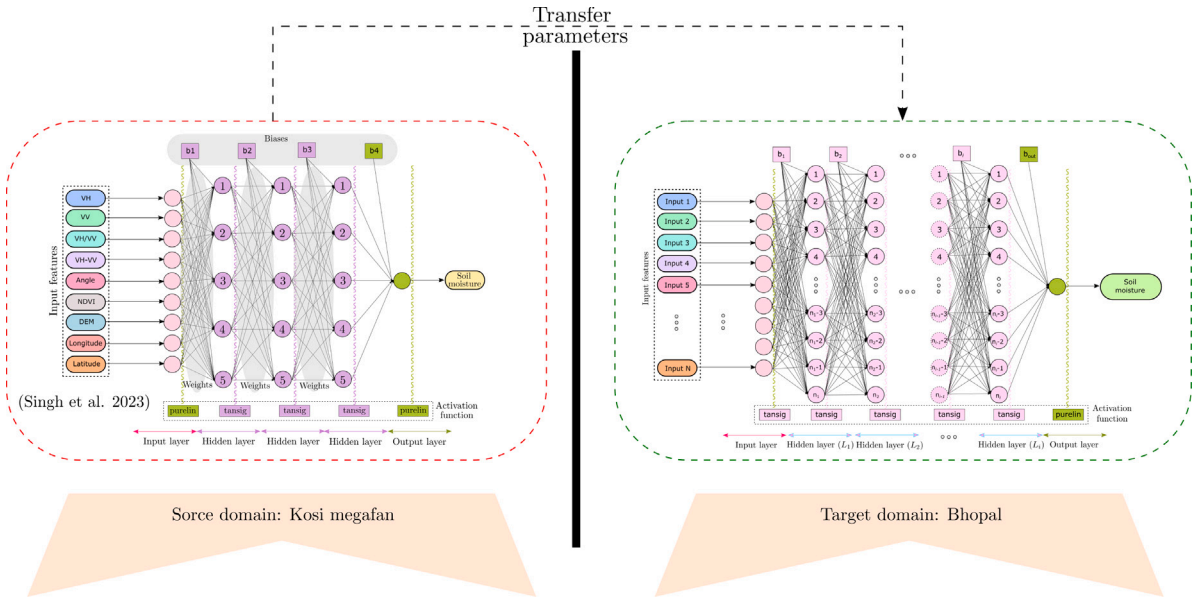

Fig. 4. Schematic representation of the fined-tuned neural network architecture in which the parameters learned from the training over a relatively large source domain (*i.e.,* the Kosi Fan) are transferred to a smaller target domain (*i.e.,* Bhopal) having sparse data for the generalisation of the soil moisture estimation. A general architecture for the target domain is presented, enabling the acquisition of hyperparameters and model parameters from the source domain.

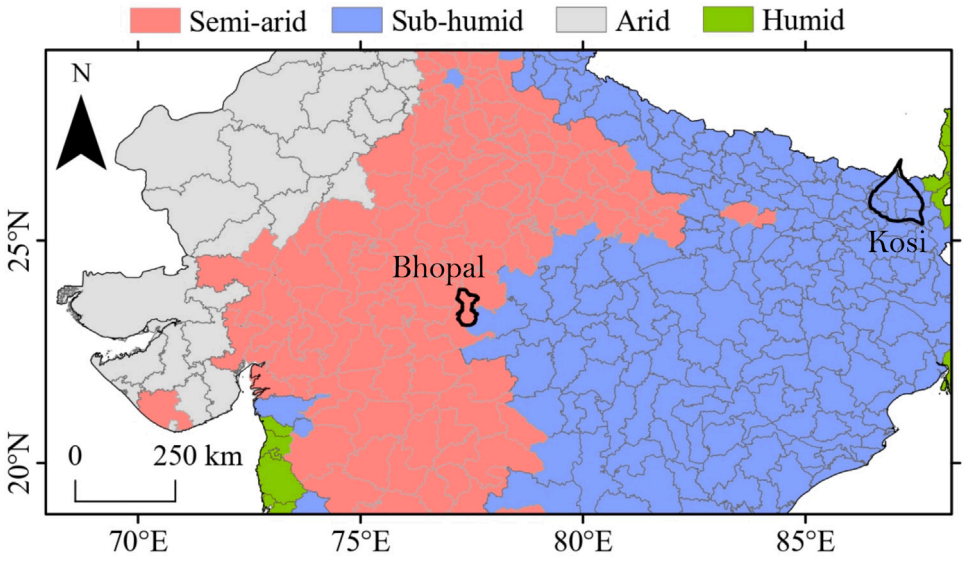

Fig. 5. Climatic zone classification of the study sites. The study sites are marked by black boundaries (Sources: ICAR-CRIDA (2013) and Singh and Gaurav (2024)).

impact of all other features. To do so, we utilise two complementary techniques: individual conditional expectation (ICE) curves, and partial dependency plots (PDP) (Goldstein et al., 2015). These techniques are frequently used to investigate the influence of the feature set on the response variable (Singh et al., 2021; Singh and Gaurav, 2023; Singh et al., 2023a). ICE curves demonstrate the effect of each feature at individual observations, while PDP provides an overall view of a feature's impact by averaging the effects of other features. The PDP for the m^{th} feature is derived using Equation (3).

$$\chi(f_m) \approx \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \chi(f_m, \boldsymbol{F} - f_m) \tag{3}$$

where N_{obs} denotes the total observations. The ICE curves are obtained by disaggregating Equation (3). A detailed explanation of the feature sensitivity technique used in this study can be found in Singh et al. (2024).

2.4. Transfer learning model

Transfer learning involves adapting the existing learning or knowledge from a source domain to a target domain (Fig. 4). It involves optimising the target domain model parameters Θ_{target}, using the source domain model parameters Θ_{source}, and additional target domain data. This can be mathematically expressed as:

$$min_{\Theta_{\text{target}}} \mathcal{L}_{\text{target}}(f(\Theta_{\text{source}})) \tag{4}$$

This equation signifies minimising the target domain loss, $\mathcal{L}_{\text{target}}(\cdot)$, by adapting the source domain parameters to the target domain through the transfer function $f(\cdot)$. In this study, we adopted the deep learning and data fusion framework proposed by Singh and Gaurav (2023), which utilises a fully connected feed-forward neural network model as the source domain model. Their framework incorporates nine satellite-derived input features: VH, VV, VH/VV, VH-VV, incidence angle, NDVI, elevation, longitude, and latitude. Through extensive training

with 224 in situ datasets, they developed a 9:5:5:5:1 neural network structure to estimate surface soil moisture at the Kosi Fan (referred to as Task A). We extracted crucial information from this model, including its hyper and model parameters, and utilised them as the initial conditions for training a separate model (referred to as Task B) aimed at estimating soil moisture in the climatically distinct region, Bhopal (Fig. 5). It is important to note that the choice of the transfer function and optimisation strategy may vary depending on the unique characteristics of both the source and target domains, as well as the nature of the available data. For Bhopal, we adopted the same 9:5:5:5:1 architecture with a minor modification in the activation function at the input layer. Instead of the linear "purelin" activation used by Singh and Gaurav (2023), we adopted the hyperbolic tangent sigmoid ("tansig") activation to setup the model for Bhopal, which introduces non-linearity to better capture soil moisture fluctuations in this semi-arid region and significantly reduces model error. Instead of initialising the model parameters (weights and biases) randomly, we leveraged the optimal parameters from Singh's model as a starting point. This new model was then further refined by fine-tuning it on the limited dataset of Bhopal (reduced by 55% as compared to the Kosi) using the Levenberg–Marquardt backpropagation algorithm (Marquardt, 1963). By leveraging the inherent learning of the pre-existing model and enhancing it through adaptation to the new location, the updated model gains a new transformative capability. This process effectively integrates insights from both domains into a unified (or joint) model. The detailed pseudocode of the proposed transfer learning framework is presented in the Algorithm 1.

Algorithm 1 Pseudo-code for the transfer learning framework.

Definitions:
- **Pre-trained Model**: A model that has been trained on a large dataset for soil moisture estimation.
-**Task A:** Soil moisture estimation at Kosi Fan (humid subtropical) with large in situ measurements
- **Source Domain**: The domain from which the pre-trained model is obtained (Task A - Kosi Fan).
- **Fine-tuning**: Optimising pre-trained model parameters for better performance on a smaller dataset.
-**Task B:** Soil moisture estimation at Bhopal (semi-arid) with limited in situ measurements
- **Target Domain**: The domain to which the pre-trained model is adapted (Task B - Bhopal).
procedure EXTRACT DETAILS (Task A model)
  Extract model architecture, hyperparameters, and model parameters
  From the 9-5-5-5-1 fully connected feed-forward neural network proposed by Singh and Gaurav (2023)
  Return optimal hyperparameters and model parameters ($\Theta_{pretrained\_A}$) and model details
end procedure
Input:
  Pre-trained Model parameters: $\Theta_{pretrained\_A}$ (weights, biases)
  Fine-tuning model parameters: $\Theta_{fine\_tune\_B}$ (weights and biases)
  Task A (Source Domain) Dataset: D_{task_A} (Kosi Fan - Humid Subtropical) ▷ Optional
  Task B (Target Domain) Dataset: D_{task_B} (Bhopal - Semi-arid)
Output:
  Fine-tuned Model for Task B: new_model_B
procedure MAIN
  Load Pre-trained Model trained on Task A
  pretrained_model_A ← load_pretrained_model_task_A ($\Theta_{pretrained\_A}$)
  Modify Pre-trained Model for Task B
  new_model_B ← modify_pretrained_model_for_task_B (pretrained_model_A)
  Freeze Layers
  freeze_layers(new_model_B) ▷ Optionally freeze certain layers
  Load New Dataset for Task B (Bhopal)
  dataset_B ← load_dataset_Bhopal(D_{task_B})
  for each model parameter set $\Theta_{fine\_tune\_B}$ **do**
    Fine-tune Model for Task B using Levenberg-Marquardt backpropagation
    fine_tune_model_B(new_model_B, dataset_B, $\Theta_{fine\_tune\_B}$)
    Evaluate Model on Task B
    evaluation_results_B ← evaluate_model_B (new_model_B, dataset_B) ▷ Evaluate performance
    Computing performance metrics (R, RMSE, bias)
  end for
  Select hyperparameters and model parameters with the best performance
  Save Trained Model for Task B
  save_model_B(new_model_B) ▷ Save fine-tuned model
end procedure
Main()

3. Results

3.1. Feature importance and sensitivity analysis

The feature importance analysis reveals significant insights into the mapping of surface soil moisture by using satellite-derived input features. It is worth noting that VV emerges as the foremost influential input feature, closely followed by latitude, incident angle, and longitude (Fig. 6a). Surprisingly, synthetic features such as VH/VV, and DEM exhibit greater relevance than NDVI, highlighting their crucial role in the process. Conversely, VH-VV appears to be the least significant feature. Moreover, the feature association estimate reveals an absence of correlation among input features, suggesting their independent contributions to mapping surface soil moisture (Fig. 6b). This implies that all input features can be effectively utilised in the process without redundancy or overlapping influence.

The feature sensitivity analysis, conducted via PDP plots and ICE curves, delineates the individual impact of each feature by isolating its effect from other features. We observed a dual behaviour for VH (Fig. 7). Specifically, certain ICE curves decrease (indicating high VH values for low soil moisture), while others increase (suggesting high VH values for high soil moisture) between 0.015–0.02. The former points belong to measurements taken from vegetated regions, whereas the latter belong to measurements taken from barren land. A clear increasing trend is observed for VV. Physically, C-band VV backscatter is highly sensitive to the soil's dielectric constant, which increases sharply as volumetric water content rises. In wetter soils, the stronger dielectric contrast at the air-soil interface and within the pore space causes more energy to be scattered back towards the sensor in VV polarisation. Thus, the PDP's positive trend, showing higher predicted moisture with increasing VV, is fully consistent with established scattering models (e.g., IEM) and numerous experimental studies that link rising VV backscatter to higher soil moisture (Singh and Gaurav, 2023). However, for both synthetic features, no distinct trend is evident except for a slight undulating increase for VH/VV. As for the incident angle, no clear trend is observed. NDVI and DEM exhibit a slight increasing trend. Furthermore, we observe a clear increasing and decreasing trend with latitude and longitude, respectively.

3.2. Performance of the TL-SSM

Once the model is tuned for the target domain (Bhopal), we evaluate its training accuracy using the TL-SSM model. We feed the training data into the model and recorded the predicted response. We observed a strong agreement between predicted and observed values, with an R value of 0.83, an RMSE of 0.05 m^3/m^3, and zero bias (Fig. 8a).

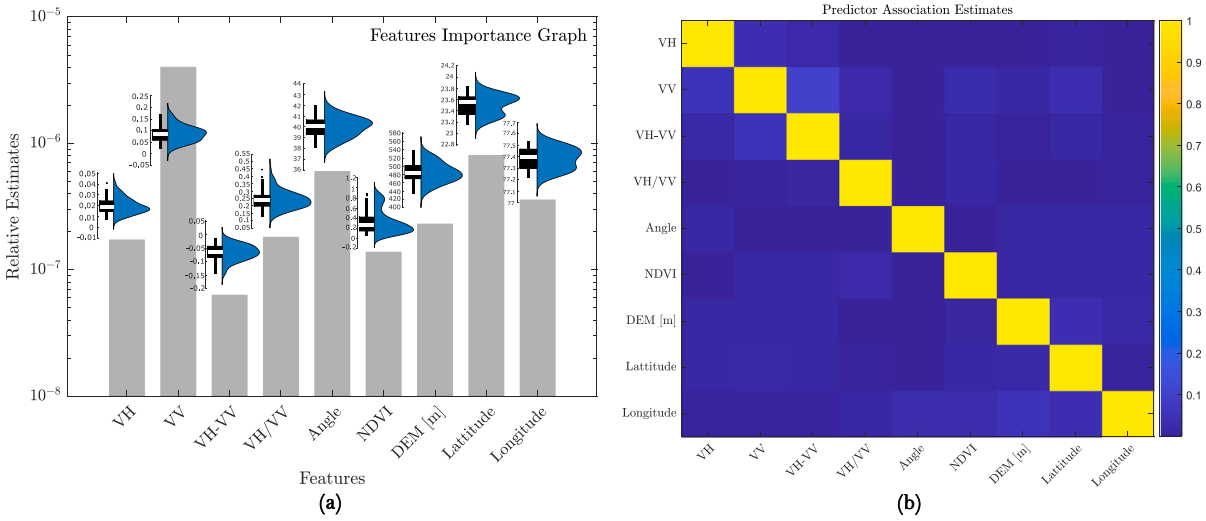

Fig. 6. (a) Feature importance score of the input feature with corresponding box and half-violin plot. NDVI, latitude, and longitude data follow a bimodal distribution, whereas the rest of the input features follow a Gaussian distribution. (b) Feature association estimate.

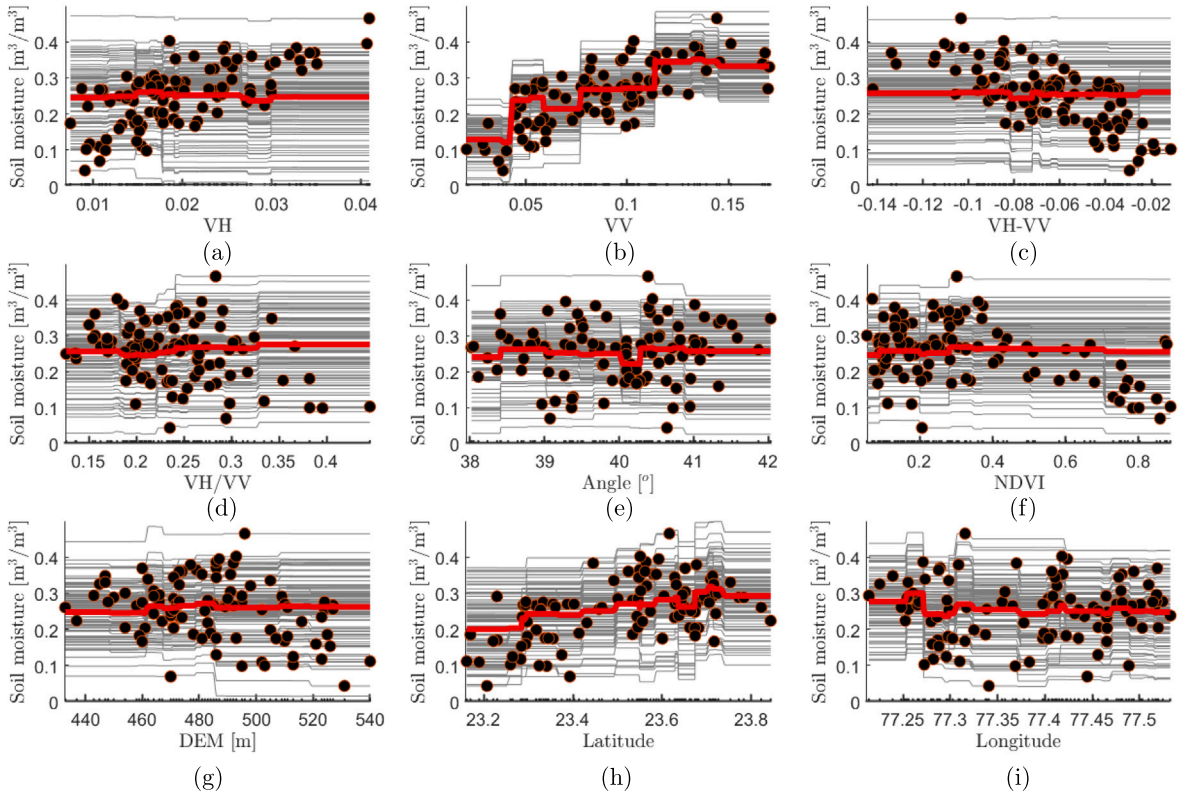

Fig. 7. Plot (a)–(i) illustrates the feature sensitivity analysis of all the input features. The black dot indicates the observation points. The grey lines indicate the individual conditional expectation curve, and the thick red line represents the partial dependence curve.

However, high training accuracy does not necessarily ensure high accuracy on unseen datasets, as it may result from overfitting to local fluctuations in the training data. To mitigate overfitting, we assessed the model's performance on validation data while adjusting model parameters. During validation, we noticed a strong alignment between predicted and observed data, with an R value of 0.89, an RMSE of 0.03 m^3/m^3, and a bias of 0.02 m^3/m^3 (Fig. 8b). A small positive bias suggests a slight overestimation of soil moisture by the model. Subsequently, we evaluated the model's performance on unseen testing data, finding reasonably good results with an R value of 0.81, an RMSE of 0.053 m^3/m^3, and a bias of 0.023 m^3/m^3, comparable to the training accuracy (Fig. 8c). Finally, we reported an overall accuracy (R = 0.82, RMSE = 0.05 m^3/m^3, and bias = 0.01 m^3/m^3) by evaluating the model's performance across the entire dataset, encompassing training, validation, and testing sets (Fig. 8d).

To quantify the advantage of our transfer learning framework, we conducted a control experiment. We applied the initial model directly to the target domain without transferring model parameters from the source domain (*i.e.,* with random initial parameters). We used the same data split (60:10:30 for training, validation, and testing) for both scenarios. This baseline model achieved a lower testing accuracy (R = 0.69, RMSE = 0.0593, and bias = 0.03). This result highlights the

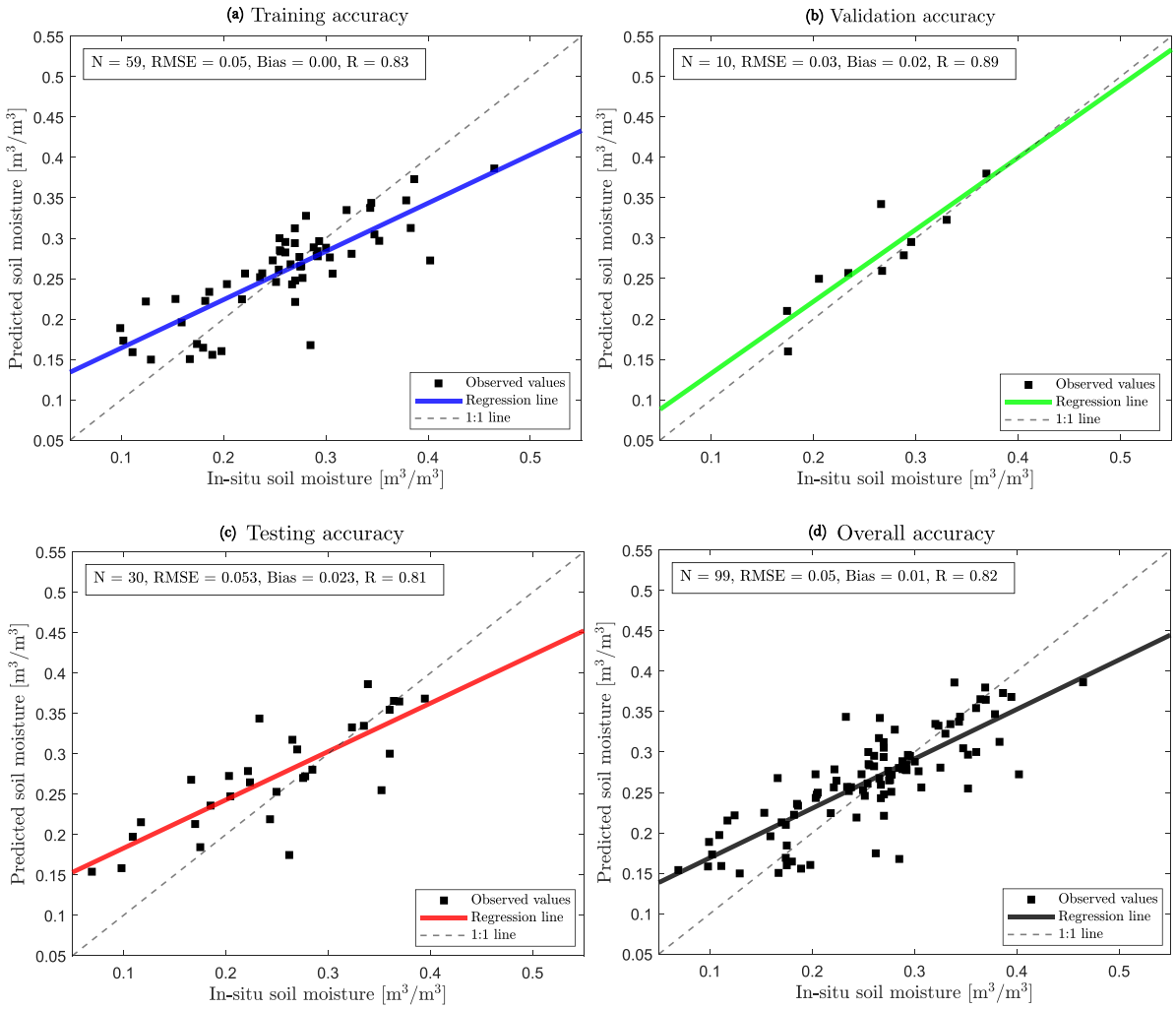

Fig. 8. Performance of the transfer learning model on the training, validation, testing, and overall datasets (a–d). The dashed line indicates the y=x line.

effectiveness of our proposed framework, which achieves improved accuracy on the target domain with 55% less in situ data compared to the source domain.

3.3. Error histogram and residual analysis

We performed error histogram analysis to examine the distribution of model errors. We computed the error (*i.e.,* observed – predicted) for the training, validation, and testing phases, and plotted the error histogram using an 11-bin size to denote symmetry (Fig. 9). We observed that the error histogram ranged from −0.099 (in the leftmost bin) in the overestimation region to 0.118 (in the rightmost bin) in the underestimation region. Subsequently, we fitted a Gaussian distribution to the error histogram and discovered that the peak of the Gaussian curve precisely aligned with the zero error line. This alignment suggests that the majority of errors are centred around zero, indicating a good fit model.

Further, we perform residual analysis to assess the robustness and reliability of the proposed model (Fig. 10). It enables us to identify outliers that could potentially disrupt the model's performance. Additionally, it assists in detecting heteroscedasticity, which occurs when the residuals exhibit varying variances across different levels of the variables. Residual plots serve as a tool for spotting patterns indicative of heteroscedasticity. Furthermore, by conducting residual analysis, we can ascertain whether the model adequately captures nonlinear relationships between the input and response variables. Our observations reveal that the residuals display stochastic behaviour without any discernible pattern, suggesting that no further adjustments are required to improve the model's predictive accuracy.

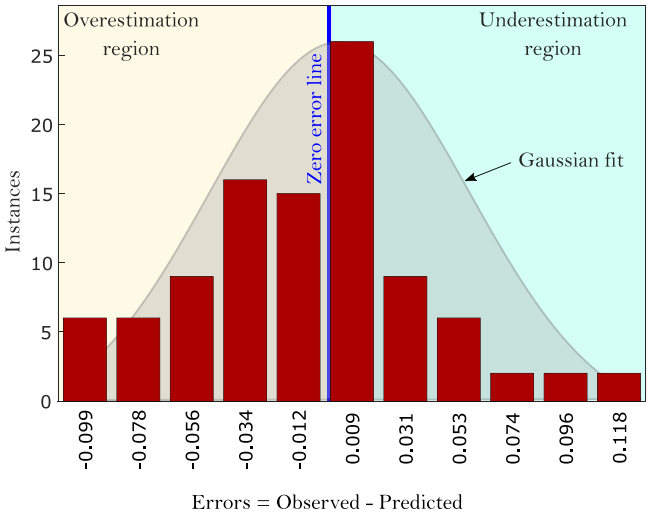

Fig. 9. The bar graph depicts the error histogram analysis, with the vertical blue line representing the zero error threshold. Areas to the left and right of this line correspond to overestimation and underestimation regions, respectively.

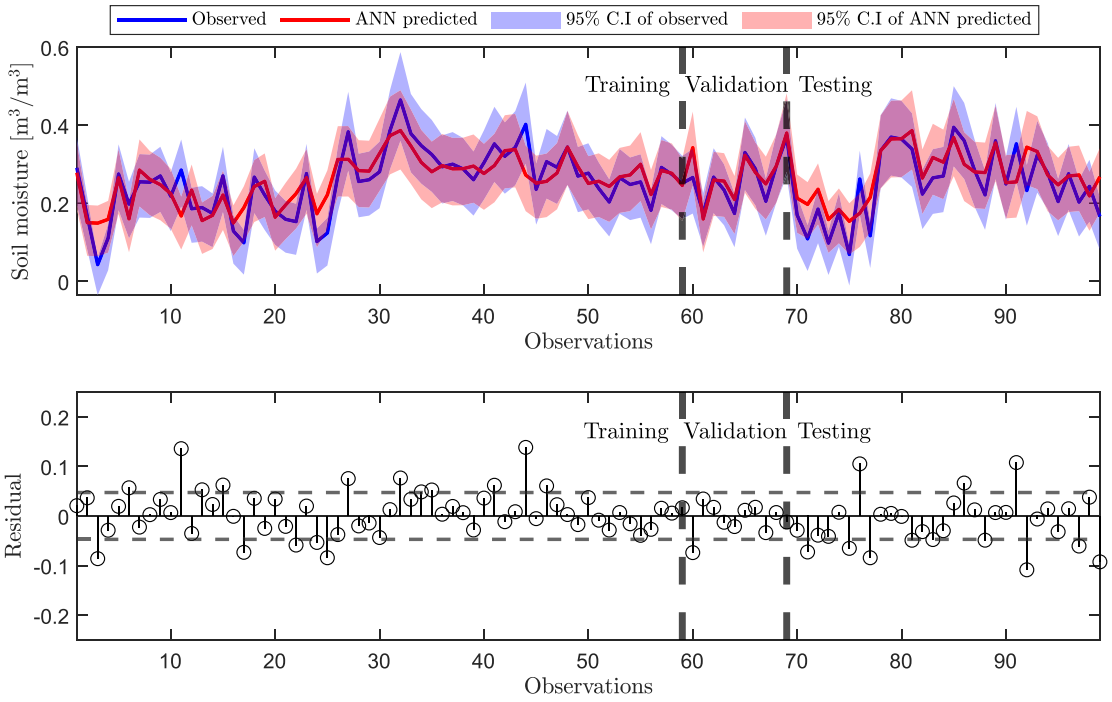

Fig. 10. The line plot in the upper panel shows the observed versus predicted soil moisture across the training, validation, and testing phases. The lower panel displays the residuals between observed and fitted values, with the dashed line representing the overall RMSE values.

4. Discussion

4.1. Comparison with benchmark algorithms

For a robust and unbiased evaluation, we compared our transfer-learning model against ten well-established benchmarks chosen to span diverse learning paradigms, computational costs, and proven utility in remote sensing and hydrological applications. These include kernel methods (Gaussian Process Regression (GPR) Williams and Rasmussen, 1995, Support Vector Regression (SVR) Drucker et al., 1996), neural networks (Generalised Regression Neural Network (GRNN) Specht et al., 1991, Radial Basis Neural Network (RBNN) Lowe, 1989, Exact RBN (ERBN) Broomhead and Lowe, 1988), tree-based models (Binary Decision Tree (BDT) Laurent and Rivest, 1976, Random Forest (RF) Breiman et al., 1984), additive models (Generalised Additive Model (GAM) Hastie and Tibshirani, 1987), ensemble learning (Boosting EL Zhang and Ma, 2012), and Automated Machine Learning (AutoML He et al., 2018). We selected these algorithms for their complementary strengths, ranging from interpretable low-cost models to high-capacity learners, and their widespread validation in soil moisture and environmental studies (Singh et al., 2023c; Nagar et al., 2023; Singh and Gaurav, 2023; Singh et al., 2024).

In addition to traditional metrics such as R, RMSE, and bias, we have incorporated advanced criteria, including Akaike Information Criterion (AIC), Corrected AIC (AICc), and Bayesian Information Criterion (BIC) for a more effective multi-model comparison (Vrieze, 2012; Claeskens and Hjort, 2008). These metrics penalise the model that has a high number of model parameters, hence the model with the lower value is preferred. Each benchmark algorithm was trained on the same training datasets and subsequently evaluated on the corresponding testing datasets. Table 2 report the performance metrics. We observe that the proposed transfer learning framework outperforms all ten benchmark algorithms with R = 0.82, RMSE = 0.05, and bias = 0.01. RF (R = 0.56, RMSE = 0.04, and bias = 0.03), GPR (R = 0.47, RMSE = 0.05, and bias = 0.05), and AutoML (R = 0.42, RMSE = 0.02, and bias = 0.04) algorithm ranks second, third, and fourth, respectively. It is worth noting that while RMSE reflects the average magnitude of prediction errors in the original units, the correlation coefficient captures only the strength of the linear relationship, meaning a model that simply predicts the mean can achieve a deceptively low RMSE yet still yield a near-zero R by failing to capture point-to-point variability. This explains why, despite achieving RMSEs of the same order of magnitude, low-R methods like GPR (R = 0.47, RMSE = 0.0470) and RF (R = 0.56, RMSE = 0.0441) fail to capture the underlying trends or match the low error rates achieved by TL-SM (R = 0.82, RMSE = 0.0480). Based on AIC, AICc, and BIC, AutoML exhibits the best results with the lowest values. ERBN, BDT, and RBNN algorithms perform the worst out of all the benchmarks with (R < 0.05).

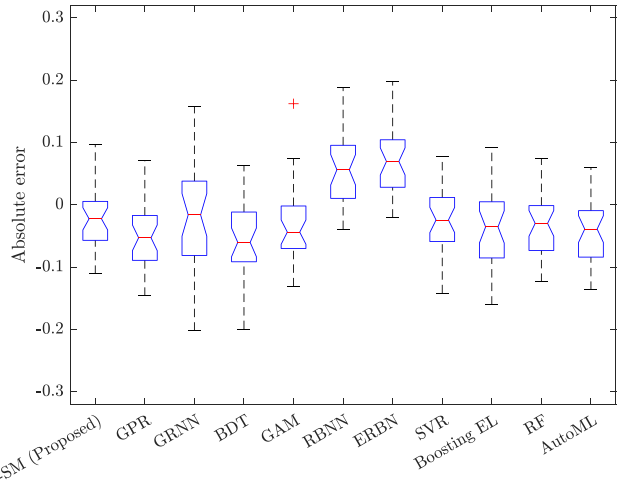

Fig. 11. Boxplot showing the analysis of model errors using ANOVA. The upper and lower edges of each box represent the 75th and 25th percentiles, respectively, while the line inside each box indicates the median of the samples.

Table 2
Comparison of the proposed model with the benchmark algorithm.

Algorithms	Proposed (TL-SM)	GPR	GRNN	BDT	GAM	RBNN	ERBN	SVR	Boosting EL (LSBoost)	RF	AutoML
R	0.82	0.47	0.06	0.02	0.29	0.02	0.01	0.19	0.20	0.56	0.48
RMSE	0.0480	0.0470	0.0531	0.0375	0.0557	0.0531	0.0531	0.0522	0.0521	0.0441	0.0166
Bias	0.01	0.05	0.02	0.06	0.03	−0.06	−0.07	0.03	0.04	0.03	0.04
AIC	191.82	−233.82	1.3581E+03	−196.54	−171.34	1.31E+03	1.29E+03	−227.77	−175.06	−234.73	−623.41
AICc	−89.94	−200.89	−192.26	−163.61	−138.41	−247.00	−253.58	−194.85	−139.41	−199.08	−552.04
BIC	354.36	−231.02	2.42E+03	−193.73	−168.53	−2.37E+03	2.36E+03	−224.97	−168.06	−227.73	−618.94

Table 3
Comparison results from the post hoc tests conducted on the models.

Group A	Group B	Lower limit	A-B	Upper limit	P-value
TL-SM	GPR	−0.022	0.027	0.076	0.789
TL-SM	GRNN	−0.047	0.002	0.051	1
TL-SM	BDT	−0.016	0.033	0.082	0.533
TL-SM	GAM	−0.042	0.007	0.056	1
TL-SM	RBNN	−0.128	−0.079	−0.030	1.12e−05
TL-SM	ERBN	−0.143	−0.094	−0.045	4.12e−08
TL-SM	SVR	−0.045	0.004	0.053	1
TL-SM	Boosting EL	−0.033	0.017	0.066	0.992
TL-SM	RF	−0.040	0.009	0.058	0.999
TL-SM	AutoML	−0.033	0.016	0.066	0.993
GPR	GRNN	−0.075	−0.025	0.024	0.854
GPR	BDT	−0.043	0.006	0.055	1
GPR	GAM	−0.070	−0.021	0.029	0.961
GPR	RBNN	−0.156	−0.107	−0.057	1.26e−10
GPR	ERBN	−0.170	−0.121	−0.072	4.88e−14
GPR	SVR	−0.072	−0.023	0.026	0.917
GPR	Boosting EL	−0.060	−0.011	0.038	0.999
GPR	RF	−0.067	−0.018	0.031	0.985
GPR	AutoML	−0.060	−0.011	0.038	0.999
GRNN	BDT	−0.018	0.031	0.080	0.620
GRNN	GAM	−0.044	0.005	0.054	1
GRNN	RBNN	−0.130	−0.081	−0.032	5.85e−06
GRNN	ERBN	−0.145	−0.096	−0.047	1.88e−08
GRNN	SVR	−0.047	0.002	0.0515	1
GRNN	Boosting EL	−0.034	0.015	0.064	0.997
GRNN	RF	−0.042	0.007	0.057	0.999
GRNN	AutoML	−0.035	0.015	0.064	0.997
⋮	⋮	⋮	⋮	⋮	⋮

Group A	Group B	Lower Limit	A-B	Upper Limit	P-value
BDT	GAM	−0.075	−0.026	0.023	0.824
BDT	RBNN	−0.161	−0.112	−0.063	6.71e−12
BDT	ERBN	−0.176	−0.127	−0.078	1.46e−15
BDT	SVR	−0.078	−0.029	0.020	0.727
BDT	Boosting EL	−0.066	−0.016	0.033	0.992
BDT	RF	−0.073	−0.024	0.025	0.901
BDT	AutoML	−0.066	−0.017	0.033	0.991
GAM	RBNN	−0.135	−0.086	−0.037	9.63e−07
GAM	ERBN	−0.150	−0.101	−0.051	2.17e−09
GAM	SVR	−0.052	−0.003	0.047	1
GAM	Boosting EL	−0.039	0.010	0.059	0.999
GAM	RF	−0.047	0.003	0.052	1
GAM	AutoML	−0.039	0.010	0.059	0.999
RBNN	ERBN	−0.064	−0.015	0.035	0.997
RBNN	SVR	0.034	0.083	0.133	2.47e−06
RBNN	Boosting EL	0.047	0.096	0.145	1.78e−08
RBNN	RF	0.039	0.089	0.138	3.58e−07
RBNN	AutoML	0.047	0.096	0.145	1.86e−08
ERBN	SVR	0.049	0.098	0.147	6.70e−09
ERBN	Boosting EL	0.061	0.110	0.159	1.83e−11
ERBN	RF	0.054	0.103	0.152	6.65e−10
ERBN	AutoML	0.061	0.110	0.159	1.94e−11
SVR	Boosting EL	−0.037	0.012	0.061	0.999
SVR	RF	−0.044	0.005	0.054	1
SVR	AutoML	−0.037	0.012	0.061	0.999
Boosting EL	RF	−0.056	−0.007	0.042	0.999
Boosting EL	AutoML	−0.049	−0.0001	0.049	1
RF	AutoML	−0.042	0.007	0.056	1

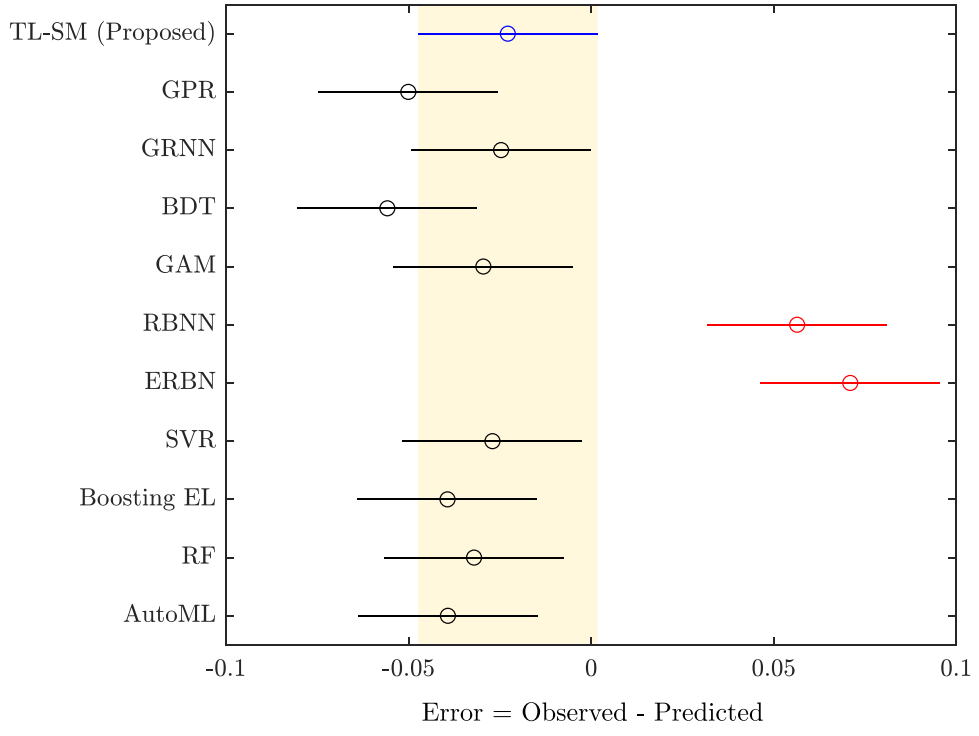

Fig. 12. Statistical comparison of the benchmark algorithm. The yellow vertical patch indicates the comparison interval of the TL-SM algorithm.

We also performed ANOVA statistical analysis to compare all the benchmark algorithms. We first calculated the errors of all ten benchmark algorithms by averaging 30 independent runs for each algorithm. We then conducted a normality test on the error distribution for each benchmark algorithm using Kolmogorov–Smirnov analysis (Smirnov, 1948; Berger and Zhou, 2014). We found that the model error follows

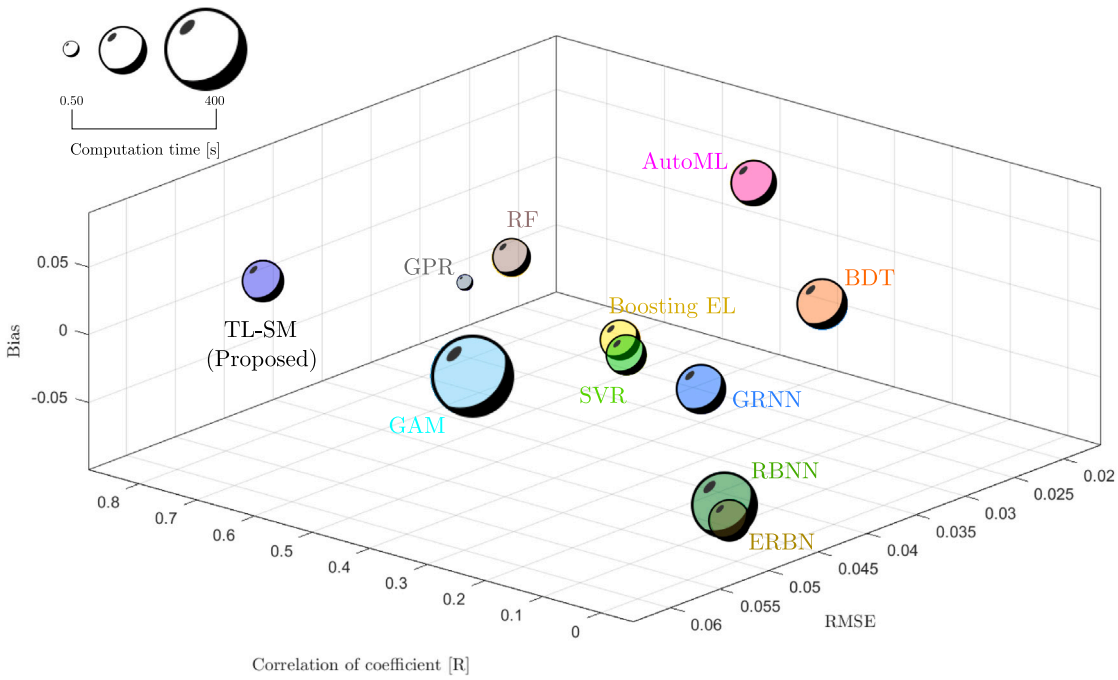

Fig. 13. Computational time alongside performance metrics (R, RMSE, and Bias) in a bubble plot. Each circle's size corresponds to the computation time of individual algorithms.

a normal distribution for all algorithms. To analyse the data, we performed one-way ANOVA with subsequent Tukey's Honestly Significant Difference (HSD) test (Abdi and Williams, 2010). This method compares the mean errors across various algorithms and compute adjusted p-values. These p-values, adjusted for multiple comparisons, provided a reliable basis for assessing the significance of differences between algorithm performances, thereby improving the precision of our evaluation. To visually analyse the distribution of model errors in the one-way ANOVA test, we plotted the box plot of model errors as shown in Fig. 11. The results of the box plot clearly support the performance metrics reported in Table 2. We found that the median error is close to zero for TL-SM (proposed), RF, GPR, and AutoML, and farthest from zero for RBNN and ERBN. For a better interpretation, we reported the pair-wise comparison of all the benchmark algorithms (Table 3). In this table, we incorporated the adjusted p-values, providing dependable indicators of the significance of the observed variations in algorithm performances. To enhance the presentation of the statistical analysis findings, we have made an attempt to illustrate the pair-wise comparison graphically (see Fig. 12). Within this figure, the blue line delineates the comparison interval of the proposed framework, extended by the yellow strip for ease of comparison. Any algorithm's comparison interval touching this yellow patch indicates that its performance is not statistically different from the proposed one (highlighted in black). Interestingly, only the performance of ERBN and RBNN significantly differs from TL-SM (highlighted in red).

Evaluating the performance of the proposed algorithm in comparison to benchmark algorithms, whether explainable or black-box, should not rely solely on performance metrics. Ignoring computational time complexity could lead to biased conclusions. For a thorough and fair assessment, we computed the computational time complexity of each algorithm using a CPU with 64 GB memory, 10 cores, operating at 3.3 GHz, and utilised a three-dimensional bubble diagram for better visualisation and interpretation of the outcomes (Fig. 13). We found that the computation time of the proposed framework, ERBN, SVR, Boosting EL, and RF is nearly the same (between 2–3 s). The time complexity of the proposed framework is represented as $\mathcal{O}(m\zeta_{l1} + \zeta_{l1} \cdot \zeta_{l2} + \cdots)$, where m denotes the number of features, and ζ_{l_i} indicates the number of neurons in layer i (Williams, 2007). Notably, GAM exhibits the highest computational time, while ERBN exhibits the second-highest.

4.2. Uncertainty analysis of the TL-SSM

We conducted a rigorous uncertainty analysis on the proposed transfer learning framework to assess its reliability under varying conditions. This analysis involved scrutinising the model's response to fluctuations in input features. Specifically, we introduced small uncertainties, ranging from ±5% to ±10%, individually into each input feature while holding all others constant. Realistic ±5-10% uncertainties in SAR backscatter and NDVI stem from speckle noise, radiometric calibration drift, atmospheric effects, DEM vertical errors, and resampling artefacts. Testing under these perturbations confirms that our transfer-learning framework remains robust to the combined impact of such satellite-derived measurement errors. The resulting variations in model output were recorded, and the percentage changes are presented in Fig. 14. Our investigation revealed that the percentage change in soil moisture spanned from −16.7% to +21.8% for ±5% and ±10% uncertainties. Among the features, VH demonstrates exceptional stability, exhibiting a marginal percentage change in soil moisture ranging from −1.4% to 2.5%. DEM followed as the second most stable feature, with a percentage change of 1.2% to 4.3%. Conversely, NDVI emerged as the most susceptible feature, displaying significant sensitivity to small uncertainties. Additionally, the synthetic feature VH-VV ranked second among the highly prone features. Crucially, VV, the most relevant feature, showcased a percentage change in soil moisture ranging from −5.4% to 1.4% under small uncertainties. The uncertainty analysis highlights the robustness of the proposed transfer learning framework, particularly with respect to the stability of certain input features. The findings provide valuable insights for optimising model performance and enhancing its reliability in practical applications. By identifying features that are susceptible to uncertainties, this study provides guidelines for refining the framework to better accommodate real-world variability.

4.3. Spatial distribution analysis

When dealing with spatial machine learning, particularly regarding surface soil moisture, it becomes crucial to evaluate the dependency of the proposed model on local factors such as topography, soil composition, vegetation cover, and land use, all of which significantly

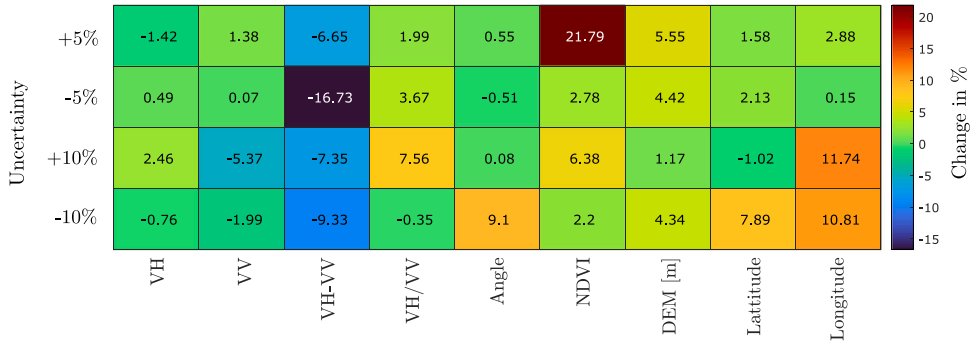

Fig. 14. Performance evaluation of the proposed model under input feature uncertainties.

influence soil moisture levels (Merz and Plate, 1997). Conducting a spatial analysis allows for the identification of these variations, facilitating the development of more precise predictive models that can accommodate spatial heterogeneity. To ensure reliable soil moisture predictions, the model should ideally exhibit spatial independence. To achieve this, we initially generated thirty independent datasets from the available data for training, validation, and testing purposes using the Mersenne Twister random generator. This involved inputting thirty distinct random seeds into the model to select in situ locations for training, validation, and testing sites, resulting in thirty unique scenarios.

We proceeded to retrain the proposed model using these thirty training datasets and assessed its performance on the corresponding validation and testing datasets. Table 4 reports the training, validation, testing, and overall accuracy for each scenario. We noticed only minor variations in the performance of the proposed model, indicating its robustness and stability. We observe the following variations in the performance metrics for training (0.81 ± 0.079, 0.05 ± 0.010, and 0.00 ± 0.008), validation (0.75 ± 0.110, 0.05 ± 0.017, and 0.01 ± 0.018), testing (0.70 ± 0.048, 0.06 ± 0.008, and −0.002 ± 0.016), and overall accuracy (0.76 ± 0.056, 0.05 ± 0.006, and 0.00 ± 0.007). The stability observed in the spatial distribution analysis suggests that the relationships between soil moisture and environmental factors remain consistent across different terrains. This highlights the reliability of the observed patterns and relationships across diverse geographic regions and topographical variations. It indicates that local factors do not substantially influence the proposed model, thus affirming its applicability across various geographical or geometric conditions. Furthermore, this stability not only reinforces the concept of the transfer learning model but also improves its interpretability. Consequently, stakeholders can make informed decisions based on dependable predictive insights.

4.4. Ablation analysis

We conducted an ablation analysis to thoroughly examine the contribution of individual components within the model. This analytical approach aids in identifying the essential driving elements of the model. In this process, we generated various scenarios by altering input features and model architecture. In the input feature ablation, we explored eight combinations of input features based on their relevance in predicting soil moisture. For model architecture ablation, we investigated six different scenarios involving the selective removal or alteration of hidden layers and the number of neurons within each layer. Across these fourteen scenarios, we recorded training, validation, testing, and overall accuracy using R, RMSE, and bias as performance metrics (Table 5). Subsequently, we assigned performance ratings for each metric (*i.e.,* R-rating, RMSE-rating, and bias-rating) and reported the overall rating. Regarding input feature ablation, our findings indicate that the proposed model yields the best results when all nine inputs are considered. Additionally, combinations involving radar backscatters (*i.e.,* {VV, VH} and {VV, VH, VH-VV, and VH/VV}) also demonstrate strong performance, with an overall rating in the 'very good' category. Concerning architecture ablation, our results reveal that the proposed model with three hidden layers, each consisting of five neurons, outperforms all other considered scenarios.

4.5. Comparison between source and target domain

We observed a slight variation in the predictive capacity of the input features when applied across two distinct climate settings. In the source domain, longitude and VV emerge as the most relevant features for mapping soil moisture. However, in the target domain, while these same two features remain most relevant, VV exhibits a higher feature importance score compared to longitude. The discrepancy likely arises from the strong influence of topography on drainage patterns in the Kosi Fan, which increases the relevance of longitude. In Bhopal, where such terrain complexity is lacking, VV becomes the more dominant feature. Additionally, the contribution of NDVI appears to be less prominent in both the source and target domains.

The individual impacts of all nine input features on soil moisture exhibit a remarkable similarity between the source and target domains. We note consistent patterns such as the dual behaviour of VH and the increasing trend of VV across both domains. Additionally, features like VV/VH, VH-VV, incidence angle, and NDVI display an undulating nature in both contexts. Moreover, DEM shows a slight increasing trend, while longitude experiences a sharp decrease in both domains. However, we observe a slight deviation concerning latitude, which decreases in the source domain but increases in the target domain. Regarding the performance of the deep learning model, we observe a consistent level of effectiveness across both domains.

5. Conclusion

This study introduces a novel transfer learning framework customise for soil moisture estimation, demonstrating its efficacy in extrapolating insights from a well-established model trained on a source domain to accurately predict soil moisture dynamics in a target domain. By leveraging the initial learning parameters of the existing model, we showcase the ability to achieve precise estimations even in scenarios with limited datasets or in regions with sparse data availability.

Our investigation highlights the effectiveness of features extracted from Sentinel-1/2 and SRTM satellite imagery in capturing the surface soil moisture dynamics. Our analysis reveals a striking consistency in the behaviour of input features across both source and target domains, enhancing the generalisation capability of our approach and facilitating its broader applicability. Despite the inherent differences in climatic settings between the source (Kosi) and target domains (Bhopal), our proposed framework consistently delivers accurate soil moisture estimations (with 55% less datasets as compared to the source

Table 4
Results of the spatial distribution analysis using thirty independent sets of training, validation, and testing datasets.

Scenarios/Seeds	Training			Validation			Testing			Overall		
	R	RMSE	Bias	R	RMSE	Bias	R	RMSE	Bias	R	RMSE	Bias
1	0.83	0.05	0	0.89	0.03	0.02	0.81	0.053	0.023	0.82	0.05	0.01
2	0.71	0.06	0.01	0.89	0.04	−0.03	0.62	0.055	−0.03	0.68	0.06	−0.01
3	0.87	0.04	0	0.81	0.05	−0.02	0.72	0.059	0	0.79	0.05	0
4	0.85	0.04	0	0.80	0.05	0.01	0.60	0.071	0.006	0.78	0.05	0.01
5	0.82	0.05	0.02	0.81	0.04	0.03	0.76	0.075	0.031	0.65	0.06	0.02
6	0.61	0.07	−0.02	0.74	0.07	0.01	0.74	0.05	−0.028	0.64	0.06	−0.02
7	0.89	0.04	0	0.62	0.07	0	0.73	0.081	−0.003	0.76	0.06	0
8	0.84	0.04	0	0.85	0.04	0.02	0.71	0.057	−0.012	0.80	0.05	0
9	0.84	0.04	0	0.98	0.02	0	0.67	0.063	−0.015	0.81	0.05	0
10	0.86	0.04	−0.01	0.71	0.09	−0.03	0.74	0.069	−0.013	0.73	0.06	−0.01
11	0.75	0.05	−0.01	0.48	0.07	0.01	0.77	0.056	0.006	0.77	0.06	0
12	0.76	0.05	0	0.70	0.05	0.03	0.65	0.069	0.025	0.69	0.06	0.01
13	0.75	0.05	0	0.73	0.05	0.01	0.7	0.063	−0.004	0.74	0.06	0
14	0.87	0.04	0	0.60	0.05	0	0.72	0.055	−0.017	0.81	0.05	0
15	0.79	0.05	0	0.69	0.03	0	0.66	0.073	−0.026	0.72	0.06	−0.01
16	0.89	0.04	0	0.66	0.07	0.02	0.74	0.056	−0.011	0.81	0.05	0
17	0.82	0.05	−0.01	0.76	0.07	0.01	0.71	0.053	0.002	0.77	0.05	0
18	0.90	0.04	0	0.68	0.05	0.01	0.66	0.056	0.006	0.82	0.05	0
19	0.70	0.05	0	0.69	0.08	−0.02	0.67	0.066	0.014	0.67	0.06	0
20	0.81	0.04	0	0.81	0.03	−0.02	0.75	0.069	0.002	0.78	0.05	0
21	0.87	0.04	0	0.79	0.06	−0.03	0.74	0.054	0.025	0.80	0.05	0
22	0.76	0.06	0	0.55	0.04	0.01	0.73	0.055	−0.015	0.73	0.06	0
23	0.80	0.05	0.01	0.62	0.05	0	0.64	0.058	0	0.76	0.05	0.01
24	0.93	0.03	0	0.73	0.06	0.02	0.64	0.049	0.002	0.83	0.05	0
25	0.66	0.06	0.01	0.76	0.05	0	0.76	0.057	−0.017	0.70	0.06	0
26	0.73	0.06	0.02	0.74	0.07	0.01	0.70	0.052	−0.005	0.71	0.06	0.01
27	0.84	0.04	0	0.92	0.04	0.03	0.68	0.056	0.022	0.81	0.05	0.01
28	0.92	0.03	0	0.80	0.04	0.02	0.70	0.068	−0.005	0.82	0.05	0
29	0.95	0.03	0	0.88	0.03	0	0.67	0.064	−0.003	0.85	0.04	0
30	0.78	0.05	−0.01	0.78	0.07	0.04	0.69	0.053	−0.009	0.74	0.05	0
$\mu \pm \sigma$	0.81 ± 0.079	0.05 ± 0.010	0.00 ± 0.008	0.75 ± 0.110	0.05 ± 0.017	0.01 ± 0.018	0.70 ± 0.048	0.06 ± 0.008	−0.002 ± 0.016	0.76 ± 0.056	0.05 ± 0.006	0.00 ± 0.007

Table 5
Results of the input and architecture ablation analysis.

Ablation	Scenarios	Training			Validation			Testing			Overall			Overall rating
		R	RMSE	Bias	R	RMSE	Bias	R	RMSE	Bias	R	RMSE	Bias	
Input features	VH	0.75	0.06	0.01	0.67	0.07	−0.01	0.48	0.064	−0.018	0.64	0.06	0	Satisfactory
	VV	0.25	0.08	0	0.75	0.06	−0.03	0	0.073	−0.015	0.15	0.08	−0.01	Poor
	VV, VH	0.83	0.05	0	0.73	0.04	−0.01	0.59	0.065	0	0.77	0.05	0	Very Good
	VH-VV, VH/VV	0.77	0.05	0.01	0.78	0.04	−0.01	0.61	0.063	0.007	0.73	0.06	0.01	Good
	VV, VH, VH-VV, VH/VV	0.75	0.06	0	0.79	0.05	−0.01	0.78	0.053	0.002	0.75	0.05	0	Very Good
	VV, VH, VH-VV, VH/VV, angle	0.72	0.06	0.02	0.71	0.06	0.01	0.76	0.05	0.024	0.73	0.06	0.02	Good
	VV, VH, VH-VV, VH/VV, angle, NDVI	0.78	0.05	0	0.77	0.07	0.02	0.63	0.065	0.003	0.73	0.06	0	Good
	All features	0.83	0.05	0	0.89	0.03	0.02	0.81	0.053	0.023	0.82	0.05	0.01	Very Good
Architecture	9:1:1	0.65	0.07	0.03	0.79	0.03	0.03	0.7	0.059	0.048	0.67	0.06	0.04	Satisfactory
	9:5:1	0.59	0.07	−0.02	0.75	0.04	0.02	0.66	0.064	−0.021	0.61	0.06	−0.01	Satisfactory
	9:1:1:1	0.66	0.06	0	0.66	0.07	0	0.68	0.07	0.015	0.67	0.06	0.01	Satisfactory
	9:5:5:1	0.81	0.05	0	0.82	0.05	0	0.71	0.064	−0.001	0.77	0.05	0	Very Good
	9:1:1:1:1	0.46	0.07	−0.05	0.15	0.08	0.03	0.38	0.08	−0.034	0.44	0.07	−0.04	Poor
	9:5:5:5:1	0.83	0.05	0	0.89	0.03	0.02	0.81	0.053	0.023	0.82	0.05	0.01	Very Good

domain), emphasising its adaptability across diverse landscapes. This resilience demonstrate the robustness and reliability of our methodology, offering invaluable insights for diverse applications ranging from agricultural management to environmental monitoring. Our study not only significantly improves soil moisture estimation but also demonstrates the potential of transfer learning to overcome data scarcity challenges in environmental domains, enabling more efficient and accurate predictive modelling in diverse environmental contexts.

CRediT authorship contribution statement

Abhilash Singh: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **M. Niranjannaik:** Validation, Investigation, Data curation. **Kumar Gaurav:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Investigation.

Funding

This study was conducted without financial support from any public, commercial, or non-profit funding agencies.

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.

Acknowledgements

We appreciate the institutional assistance provided by IISER Bhopal. The authors sincerely appreciate the thoughtful feedback and helpful suggestions provided by the editor and the four anonymous reviewers.

Data and code availability

The input features used in this study are obtained from publicly accessible databases and can be downloaded from the respective sources, as detailed in Section 2. The corresponding code is available at https://abhilashsingh.net/codes.html.

References

Abdi, H., Williams, L.J., 2010. Tukey's honestly significant difference (hsd) test. Encycl. Res. Des. 3, 1–5.

Attema, E., Ulaby, F.T., 1978. Vegetation modeled as a water cloud. Radio Sci. 13, 357–364.

Autret, M., Bernard, R., Vidal-Madjar, D., 1989. Theoretical study of the sensitivity of the microwave backscattering coefficient to the soil surface parameters. Remote. Sens. 10, 171–179.

Baghdadi, N., Gherboudj, I., Zribi, M., Sahebi, M., King, C., Bonn, F., 2004. Semi-empirical calibration of the iem backscattering model using radar images and moisture and roughness field measurements. Int. J. Remote Sens. 25, 3593–3623.

Baier, A., Boukhers, Z., Staab, S., 2021. Hybrid physics and deep learning model for interpretable vehicle state prediction. arXiv preprint arXiv:2103.06727.

Berger, V.W., Zhou, Y., 2014. Kolmogorov–Smirnov test: Overview. Wiley statsref: Statistics reference online.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.

Broomhead, D.S., Lowe, D., 1988. Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks. 25, Royal signals and radar establishment Malvern (United Kingdom), pp. 1–8.

Claeskens, G., Hjort, N.L., 2008. Model Selection and Model Averaging. Cambridge books.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9.

Dubois, P.C., Zyl, J.Van., Engman, T., 1995. Measuring soil moisture with imaging radars. IEEE Trans. Geosci. Remote Sens. 33, 915–926.

Efremova, N., Seddik, M.E.A., Erten, E., 2021. Soil moisture estimation using sentinel-1/-2 imagery coupled with cyclegan for time-series gap filing. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Farrar, T., Nicholson, S., Lare, A., 1994. The influence of soil type on the relationships between ndvi, rainfall, and soil moisture in semiarid botswana. ii. ndvi response to soil oisture. Remote. Sens. Environ. 50, 121–133.

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Biswas, A., 2020. Comparison of spectral and spatial-based approaches for mapping the local variation of soil moisture in a semi-arid mountainous area. Sci. Total Environ. 724, 138319.

Felegari, S., Sharifi, A., Moravej, K., Golchin, A., Tariq, A., 2022. Investigation of the relationship between ndvi index, soil moisture, and precipitation data using satellite images. Sustain. Agric. Syst. Technol. 31, 4–325.

Fung, A.K., Li, Z., Chen, K.S., 1992. Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sens. 30, 356–369.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Statist. 24, 44–65.

Greifeneder, F., Notarnicola, C., Hahn, S., Vreugdenhil, M., Reimer, C., Santi, E., Paloscia, S., Wagner, W., 2018. The added value of the vh/vv polarization-ratio for global soil moisture estimations from scatterometer data. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 11, 3668–3679.

Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. J. Amer. Statist. Assoc. 82, 371–386.

He, Y., Lin, J., Liu, Z., Wang, H., Li, L.J., Han, S., 2018. Amc: Automl for model compression and acceleration on mobile devices. In: Proceedings of the European Conference on Computer Vision. (ECCV), pp. 784–800.

He, Z., Sun, Y., Li, R., 2024. Transfusion: Covariate-shift robust transfer learning for high-dimensional regression. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp. 703–711.

Hemmati, E., Sahebi, M.R., 2024. Surface soil moisture retrieval based on transfer learning using sar data on a local scale. Int. J. Remote Sens. 45, 2374–2406.

ICAR-CRIDA, 2013. Central Research Institute for Dryland Agriculture, Hyderabad, India. Annual Report, p. 182.

Ittner, A., Schlosser, M., 1996. Discovery of relevant new features by generating non-linear decision trees. In: KDD. pp. 108–113.

Joseph, A., van der Velde, R., O'neill, P., Lang, R., Gish, T., 2010. Effects of corn on c-and l-band radar backscatter: A correction method for soil moisture retrieval. Remote Sens. Environ. 114, 2417–2430.

Kumar, A., Gaurav, K., Singh, A., Yaseen, Z.M., 2024. Assessment of machine learning models to predict daily streamflow in a semiarid river catchment. Neural Comput. Appl. 36, 13087–13106.

Laurent, H., Rivest, R.L., 1976. Constructing optimal binary decision trees is np-complete. Inform. Process. Lett. 5, 15–17.

Li, Q., Wang, Z., Shangguan, W., Li, L., Yao, Y., Yu, F., 2021. Improved daily smap satellite soil moisture prediction over china using deep learning model with transfer learning. J. Hydrol. 600, 126698.

Lin, H., Kogelmann, W., Walker, C., Bruns, M., 2006. Soil moisture patterns in a forested catchment: A hydropedological perspective. Geoderma 131, 345–368.

Lowe, D., 1989. Adaptive radial basis function nonlinearities, and the problem of generalisation. In: 1989 First IEE International Conference on Artificial Neural Networks, (Conf. Publ. No. 313), IET. pp. 171–175.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11, 431–441.

Matsumoto, M., Nishimura, T., 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. (TOMACS) 8, 3–30.

McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017. The global distribution and dynamics of surface soil moisture. Nat. Geosci. 10, 100–104.

Merz, B., Plate, E.J., 1997. An analysis of the effects of spatial variability of soil and soil moisture on runoff. Water Resour. Res. 33, 2909–2922.

Mirsoleimani, H.R., Sahebi, M.R., Baghdadi, N., El Hajj, M., 2019. Bare soil surface moisture retrieval from sentinel-1 sar data based on the calibrated iem and dubois models using neural networks. Sensors 19 (3209).

Murphy, P., Ogilvie, J., Arp, P., 2009. Topographic modelling of soil moisture conditions: a comparison and verification of two models. Eur. J. Soil Sci. 60, 94–109.

Nagar, J., Chaturvedi, S.K., Soh, S., Singh, A., 2023. A machine learning approach to predict the k-coverage probability of wireless multihop networks considering boundary and shadowing effects. Expert Syst. Appl. 226, 120160.

Oh, Y., 2004. Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 42, 596–601.

Oh, Y., Sarabandi, K., Ulaby, F.T., 2002. Semi-empirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 40, 1348–1355.

Oh, Y., Sarabandi, K., Ulaby, F.T., et al., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 30, 370–381.

Ryan, J., Hubbard, A., Irvine-Fynn, T.D., Doyle, S.H., Cook, J., Stibal, M., Box, J., 2017. How robust are in situ observations for validating satellite-derived albedo over the dark zone of the greenland ice sheet?. Geophys. Res. Lett. 44, 6218–6225.

Sahebi, M., Angles, J., 2010. An inversion method based on multi-angular approaches for estimating bare soil surface parameters from RADARSAT-1. Hydrol. Earth Syst. Sci. 14 (2355).

Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., 2016. Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors. Int. J. Appl. Earth Obs. Geoinf. 48, 61–73.

Shen, Q., Wang, H., Shum, C., Jiang, L., Yang, B., Zhang, C., Dong, J., Gao, F., Lai, W., Liu, T., 2023. Soil moisture retrieval from multi-polarization sar data and potential hydrological application. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens..

Singh, A., Amutha, J., Nagar, J., Sharma, S., 2023a. A deep learning approach to predict the number of k-barriers for intrusion detection over a circular region using wireless sensor networks. Expert Syst. Appl. 211, 118588.

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

Singh, A., Gaurav, K., 2024. Piml-sm: Physics-informed machine learning to estimate surface soil moisture from multi-sensor satellite images by leveraging swarm intelligence. IEEE Trans. Geosci. Remote Sens..

Singh, A., Gaurav, K., Meena, G.K., Kumar, S., 2020. Estimation of soil moisture applying modified dubois model to sentinel-1; a regional study from central india. Remote. Sens. 12 (2266).

Singh, A., Gaurav, K., Rai, A.K., Beg, Z., 2021. Machine learning to estimate surface roughness from satellite images. Remote. Sens. 13 (3794).

Singh, A., Gaurav, K., Sonkar, G.K., Lee, C.C., 2023b.. Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: review, bibliometric analysis, applications, research findings, and future directions. IEEE Access.

Singh, A., Meena, G.K., Kumar, S., Gaurav, K., 2018. Analysis of the effect of incidence angle and moisture content on the penetration depth of l-and s-band sar signals into the ground surface. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. 4, 197–202.

Singh, A., Meena, G.K., Kumar, S., Gaurav, K., 2019. Evaluation of the penetration depth of l-and s-band (nisar mission) microwave sar signals into ground. In: 2019 URSI Asia-Pacific Radio Science Conference (AP-RASC). IEEE, p. 1.

Singh, A., Nagar, J., Amutha, J., Sharma, S., 2023c. P2ca-gam-id: Coupling of probabilistic principal components analysis with generalised additive model to predict the k- barriers for intrusion detection. Eng. Appl. Artif. Intell. 126, 107137.

Singh, A., Naik, M.N., Gaurav, K., 2022. Drainage congestion due to road network on the kosi alluvial fan, himalayan foreland. Int. J. Appl. Earth Obs. Geoinf. 112, 102892.

Singh, A., Patel, S., Bhadani, V., Kumar, V., Gaurav, K., 2024. Automl-gwl: Automated machine learning model for the prediction of groundwater level. Eng. Appl. Artif. Intell. 127, 107405.

Smirnov, N., 1948. Table for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19, 279–281.

Specht, D.F., et al., 1991. A general regression neural network. IEEE Trans. Neural Netw. 2, 568–576.

Tenenbaum, D., Band, L., Kenworthy, S., Tague, C., 2006. Analysis of soil moisture patterns in forested and suburban catchments in baltimore, maryland, using high-resolution photogrammetric and lidar digital elevation datasets. Hydrol. Process.: An Int. J. 20, 219–240.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). Psychol. Methods 17 (228).

Williams, R., 2007. Matrix–vector multiplication in sub-quadratic time:(some preprocessing required). In: SODA. pp. 995–1001.

Williams, C., Rasmussen, C., 1995. Gaussian processes for regression. Adv. Neural Inf. Process. Syst. 8.

Zhang, C., Ma, Y., 2012. Ensemble Machine Learning: Methods and Applications. Springer.

Zhu, L., Dai, J., Liu, Y., Yuan, S., Qin, T., Walker, J.P., 2024. A cross-resolution transfer learning approach for soil moisture retrieval from sentinel-1 using limited training samples. Remote Sens. Environ. 301, 113944.

Zhu, L., Si, R., Shen, X., Walker, J.P., 2022. An advanced change detection method for time-series soil moisture retrieval from sentinel-1. Remote Sens. Environ. 279, 113137.

Zhu, L., Walker, J.P., Shen, X., 2020. Stochastic ensemble methods for multi-sar-mission soil moisture retrieval. Remote Sens. Environ. 251, 112099.