PIML-SM: Physics-Informed Machine Learning to Estimate Surface Soil Moisture From Multisensor Satellite Images by Leveraging Swarm Intelligence

Abhilash Singh, *Member, IEEE*, and Kumar Gaurav, *Member, IEEE*

***Abstract*— We introduce a physics-informed machine learning (PIML) algorithm based on a feed-forward neural network (FFNN) to estimate surface soil moisture from limited in situ measurements and Sentinel-1/2 satellite images on the alluvial fan of the Kosi River by leveraging radar physics. We set up a learning bias PIML by modifying the loss function of the FFNN by using the improved integral equation model (I2EM). A particle swarm optimization (PSO) algorithm is used to optimize the tuning parameters of the PIML. The effectiveness of the proposed model is compared with ten benchmark algorithms. The performance of PIML model is superior among the benchmark algorithms, achieving a correlation coefficient (R) of 0.94, a root mean square error (RMSE = 0.019 m^3/m^3), and bias = −0.03 m^3/m^3. We conclude that the PIML model can accurately estimate soil moisture solely from satellite images, achieving higher spatial and temporal resolutions, even with limited in situ observations. The findings of this study can be applied in agriculture, hydrology, flood management, and drought monitoring, particularly in data-scarce regions.**

***Index Terms*— Improved integral equation model (I2EM), neural networks, physics-informed machine learning (PIML), Sentinel-1/2, soil moisture, swarm intelligence.**

I. Introduction

SOIL moisture is a key component in the hydrological, energy, and carbon cycle that contributes to many processes occurring at the land–atmosphere interface, such as runoff, evapotranspiration, and vegetation growth [1]. Its spatiotemporal variation mediates the weather and climate variation leading to the growth and persistence of extreme events such as heatwaves, droughts, and floods. Accurate estimation of spatial and temporal variability of soil moisture is essential for various applications in Earth and environmental sciences [2]. The spatial distribution of soil moisture is generally influenced by the precipitation, evaporation, topography, geology, and other environmental factors. The complex interactions between these factors make it challenging to model the spatiotemporal variation of soil moisture accurately.

Received 25 July 2024; revised 29 September 2024; accepted 13 November 2024. Date of publication 20 November 2024; date of current version 6 December 2024. A part of this work is funded by the Space Application Centre, Indian Space Research Organisation (SAC-ISRO), under the NASA-ISRO Synthetic Aperture Radar (NISAR) mission (Grant HYD-01). *(Corresponding author: Abhilash Singh.)*

Abhilash Singh is with the FGRS Laboratory, IISER, Bhopal 462066, India, and also with the School of Mathematics, University of Leeds, LS2 9JT Leeds, U.K. (e-mail: abhilash.singh@ieee.org).

Kumar Gaurav is with the FGRS Laboratory, IISER, Bhopal 462066, India (e-mail: kgaurav@iiserb.ac.in).

Digital Object Identifier 10.1109/TGRS.2024.3502618

It is impractical to assess soil water dynamics using field instruments at a regional or global scale. Remote sensing offers an efficient way to monitor large areas of soil moisture, overcoming the limitations of traditional methods. The two most commonly used approaches for estimating surface soil moisture are optical (i.e., shortwave and thermal radiation) and microwave (active and passive) remote sensing [3]. The underlying techniques for estimating soil moisture vary greatly for both these techniques. Most microwave-based retrieval algorithms work on the same premise, using dielectric properties of the material and how they affect reflected microwave radiation. Optical remote sensing establishes a relationship between soil moisture and image-derived parameters such as surface reflectance, variations in vegetation indices, and surface temperature. Regardless of the underlying techniques, they rely on various assumptions and approximations.

Microwave remote sensing sensors can acquire images in all weather conditions. The microwave pulses can penetrate the soil surface, enabling an estimate of moisture along the soil columns [4], [5], [6], [7], [8], [9], [10]. To simulate the permittivity value of individual pixels of quad-polarized radar images, various theoretical, empirical, and semi-empirical models have been proposed [11], [12], [13], [14], [15], [16], [17]. The pixels' permittivity values are then used as an input in the universal Topp's model to obtain the soil moisture [18]. Modifications have been proposed to use these models for dual-polarized radar images [19], [20], [21], [22], [23], [24]. However, the performance of the modified models has been observed to be sensitive to the climatic conditions. For example, Singh et al. [24], [25] noticed that the performance of the modified Dubois model to estimate soil moisture from Sentinel-1 images is different in the semi-arid and tropical humid climatic settings. This indicates that the applicability of modified backscattering models cannot be generalized across areas with different climactic conditions.

The limitation associated with the use of backscattering models triggers use of the data-driven model for estimating soil moisture [26], [27]. According to the Web of Science (WoS) database, the number of publications on soil moisture using machine learning algorithms has increased exponentially in the last few years (Fig. 1). Most of these publications mainly use random forest (RF), support vector regression (SVR), and artificial neural networks (ANNs) to predict soil moisture by using different input features derived from optical and microwave satellite images. For instance,

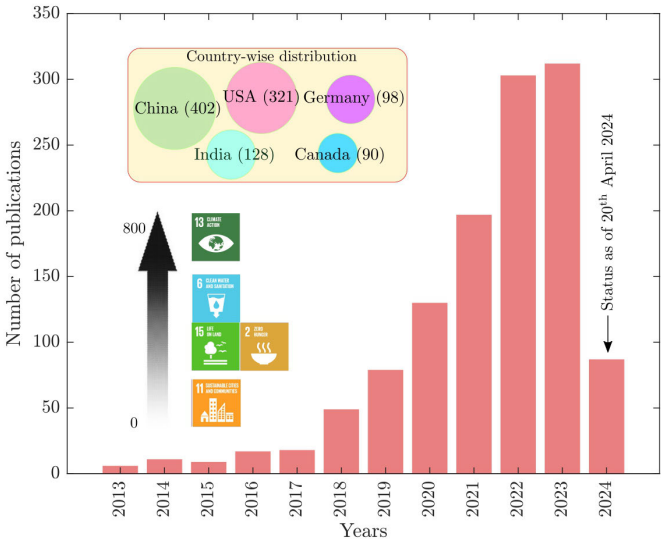

Fig. 1. Bar graph shows the number of publications on machine learning for soil moisture in the last ten years. According to the WoS database, a total of 1218 publications (research articles 1150; review articles 64, and others 4) have been published from 2013 to April 20, 2024. The maximum publications are from China, followed by the USA, India, Germany, and Canada. The majority of these publications contribute to sustainable development goals (SDGs) 13, 6, 15, 2, and 11.

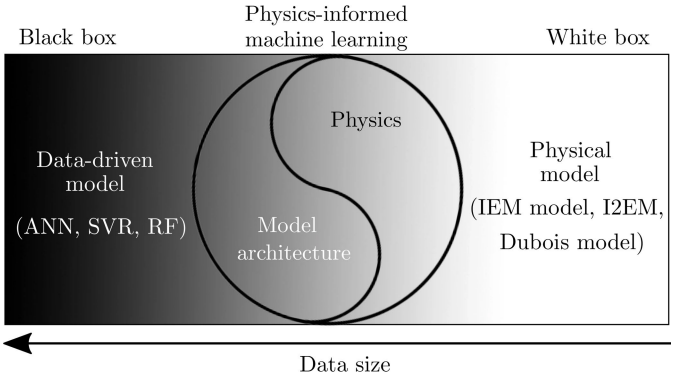

Fig. 2. Tradeoff between model interpretability and data requirements for black-box, PIML, and white-box models in soil moisture estimation.

Hajdu et al. [28] used the RF algorithm for predicting soil moisture using vertical–vertical (VV), vertical–horizontal (VH), and incidence angle information from Sentinel-1 in New Zealand. They reported a correlation coefficient of 0.93 and a root mean square error (RMSE) of 3%. Santi et al. [29] applied an ANN to develop an efficient methodology using horizontal–horizontal (HH), horizontal–vertical (HV), VV, and compact polarimetry information in Canada. They found that the model performs relatively well with R ranging from 0.70 to 0.90 and the RMSE ranges from 3% to 7%. Araya et al. [30] used unmanned aerial vehicle (UAV)-derived optical features (green, red, and NIR) to train five different machine learning algorithms [ANN, RF, SVR, relevance vector regression (RVR), and boosted regression trees (BRTs)] to estimate soil moisture. They concluded that BRT outperforms all the other algorithms with $R = 0.84$ and mean absolute error (MAE = 3.78%). Senyurek et al. [31] used the NASA's cyclone GNSS (CYGNSS) and MODIS-derived features [vegetation water content (VWC), normalized difference vegetation index (NDVI), incidence angle, surface elevation, and slope] to train three machine learning models (ANNs, RF, and SVR). They reported that RF outperforms the ANN and SVR with R = 0.90 and RMSE = 0.05 cm^3/cm^3.

Adab et al. [32] used different bands [land surface temperature (LST), blue, green, red, NIR, SWIR1, and SWIR2] of the Landsat-8 satellite to train the ANN, RF, SVR, and elastic net (EN) regression algorithms to estimate soil moisture. They reported that RF outperforms all the other algorithms with a Nash–Sutcliffe (NS) efficiency of 0.73. Cui et al. [33] applied LST and NDVI from MODIS and topographic information (surface elevation, latitude, and longitude) to train the generalized regression neural network (GRNN) model using in situ measured surface soil moisture in the Tibetan Plateau. They reported RMSE = 0.069 cm^3/cm^3 and bias = 0.003 cm^3/cm^3. Ayehu [34] considered VV, VH, and normalized difference water index (NDWI) to train the ANN and a linear regression model for soil moisture prediction in Ethiopia. They reported that the ANN performs best with R = 0.73, RMSE = 0.035 cm^3/cm^3, and bias = −0.024 cm^3/cm^3. Chaudhary et al. [35] used VV and VH polarized images of Sentinel-1 to train nine different machine learning algorithms [ANN, RF, SVR, Wang and Mendel's (WM), radial basis function (RBF), subtractive clustering (SBC), hybrid fuzzy interference system (HyFIS), and adaptive neuro-fuzzy inference system (ANFIS)] to estimate soil moisture. They reported that the SBC model outperforms all other models with R = 0.64, RMSE = 0.075 m^3/m^3, and bias = −0.009 m^3/m^3. Recently, Babaeian et al. [36] used near-infrared transformed reflectance (NTR) and NDVI information from UAV to train an automated machine learning (AutoML) model for predicting soil moisture in USA. They reported NS value of >0.9 and RMSE = 0.04 cm^3/cm^3. More recently, Singh et al. [25] used VV, VH, surface elevation, incidence angle, NDVI, latitude, and longitude information from Sentinel-1/2 and shuttle radar topography mission (SRTM) to train ANN model for soil moisture prediction. They reported R = 0.85 and RMSE = 0.05 m^3/m^3. The aforementioned studies primarily employ conventional standalone data-driven models for surface soil moisture prediction.

Regardless of high accuracy and relentless success in various scientific fields, including hydrology, the use of a data-driven model is often criticized due to the lack of physical understanding and explainability. The quality and quantity of the input features and the response variable largely govern any data-driven model's performance. Such models require a large number of input variables to predict the response variable accurately [37], [38], [39]. To overcome this problem, researchers have proposed theory-guided data science (TGDS) and physics-informed machine learning (PIML) algorithms [40], [41]. These algorithms minimize the uncertainties associated with the input features and increase the interpretability of the model. A critical investigation into the tradeoff between model interpretability and data requirements for training black-box and physics-informed

machine learning (PIML) models is paramount, especially for soil moisture prediction and its wide-ranging applications (Fig. 2).

The PIML models can be categorized as observational, inductive, and learning bias [41]. In observational bias PIML, synthetically generated features and response variables are used to train the machine learning model. The synthetic data are usually generated from simple physics rules, statistical models, or simulation techniques [42], [43], [44]. In the inductive bias PIML, prior assumptions are introduced into the model architecture so that the prediction satisfies the set of physical laws. This type of PIML has been applied in limited applications such as rotation, translation, or other symmetrical applications. The learning bias PIML uses governing physics to minimize the loss function of the machine learning model. Observational and learning bias PIML models are the most commonly used models. Numerous studies have been undertaken to investigate the potential of PIML in diverse application domains, such as defect detection and diagnosis from cardiac magnetic resonance imaging [45], [46]. For instance, Roy and Guha [47] proposed a learning bias PIML model for the solution of elastoplastic solid mechanics. They combined the residual of the governing partial differential equation (PDE), flow rules, consistency constraints, constitutive relations, and numerous boundary conditions to construct a multiobjective loss function. This study uses a learning bias PIML model by leveraging feed-forward neural networks (FFNNs) and a theoretical radar backscattering model (I2EM) to predict surface soil moisture on the Kosi Fan in the Himalayan Foreland from satellite images. Finally, we have coupled PIML with swarm intelligence for the effective optimization of the model parameters. To the best of our knowledge, PIML and swarm intelligence have not been used for soil moisture modeling. Leveraging the proposed framework, we sought to obtain empirical answers to the following research questions.

1) Can synthetic aperture radar (SAR) backscattering physics principles be incorporated into a black-box machine learning model using swarm intelligence to estimate moisture?
2) To what extent PIML models can improve the accuracy of satellite-derived soil moisture estimates compared to traditional methods?
3) Are there any tradeoffs between the interpretability of a machine learning model and the number of input variables used for training the model?

This manuscript is divided into four key sections. Section II details the input satellite and in situ observations used. It also discusses the feature engineering techniques employed to prepare the data for model training. Section III focuses on the development and optimization of the model used for analysis. It describes the model architecture, training process, and optimization techniques employed to achieve optimal performance. Section IV presents the key findings of the study. It discusses the model's performance, analyzes the results, and interprets their significance in the context of the research question. Section V summarizes the main findings of the research, reiterates the study's contributions, and discusses potential future research directions.

TABLE I
SATELLITE IMAGES USED IN THIS STUDY

Sentinel-1			Sentinel-2		
Date (DD-MM-YYYY)	**Incidence angle**	**Pass**	**Date (DD-MM-YYYY)**	**Bands**	**Resolution (spatial)**
11-12-2019	38.6	Desc.	09-12-2019	B-4 and B-8	10 m
15-12-2019	38.5	Asc.	**SRTM**		
17-12-2019	38.4	Desc.			
18-12-2019	38.5	-do-	**Date (DD-MM-YYYY)**	**Parameter**	**Resolution (spatial)**
20-12-2019	-do-	Asc.	23-09-2014	Surface elevation	30 m

II. SATELLITE IMAGES AND IN SITU OBSERVATIONS

We utilized Sentinel-1/2 images, along with the SRTM DEM, to construct a PIML model for the prediction of surface soil moisture (refer to Table I). Sentinel-1 (microwave) and Sentinel-2 (optical) are satellites for Earth observation under the Copernicus program, a collaborative effort of the European Union (EU) and the European Space Agency (ESA). They grant free and open data access through the Copernicus Open Access Hub (https://scihub.copernicus.eu/). The Sentinel-1A/B (launched in 2014 and 2016, respectively) satellites carry a C-band SAR instrument, operating at a frequency of 5.405 GHz. Sentinel-1 satellites acquire dual-polarized SAR images (like VV + VH or HH + HV). HH + HV polarization is typically used for polar and sea–ice regions. For the remaining regions, Sentinel-1 images are available in VV + VH polarization with a spatial resolution of 10 × 10 m and a swath width of 250 m. We used ground range detected (GRD) data with VV + VH polarization for our study. These raw images were processed using the sentinel application platform (SNAP) version 8.0. The processing steps includes radiometric correction to calibrate the pixel values, multilooking to reduce speckle noise (by averaging groups of pixels), speckle noise filtering with a refined Lee filter for further noise reduction, and terrain correction to account for topographic distortions. The final processed backscatter image has a resolution of 60 m × 60 m.

We also used Sentinel-2 (Level-2A) images. These images acquire data in 13 spectral bands with varying spatial resolutions ranging from 10 to 60 m. When considering both Sentinel-2A and Sentinel-2B satellites, the revisit time is five days (ten days for each individual satellite). We used the red (Band 4) and near-infrared (Band 8) bands to compute NDVI.

We downloaded the SRTM digital elevation model from the USGS EarthExplorer website (https://earthexplorer.usgs.gov/). Publicly available SRTM DEM data comes in three versions: nonvoid-filled, void-filled, and global. All versions have a spatial resolution of 30 m × 30 m (i.e., 1 × 1 arcsec). To ensure consistent spatial alignment and analysis, all data products were resampled to a common grid of 60 × 60 m using the nearest neighbor resampling technique. Following this preprocessing step, we extracted features relevant to soil moisture prediction. These features include backscatter coefficients (VH and VV polarizations) from Sentinel-1, the incidence angle

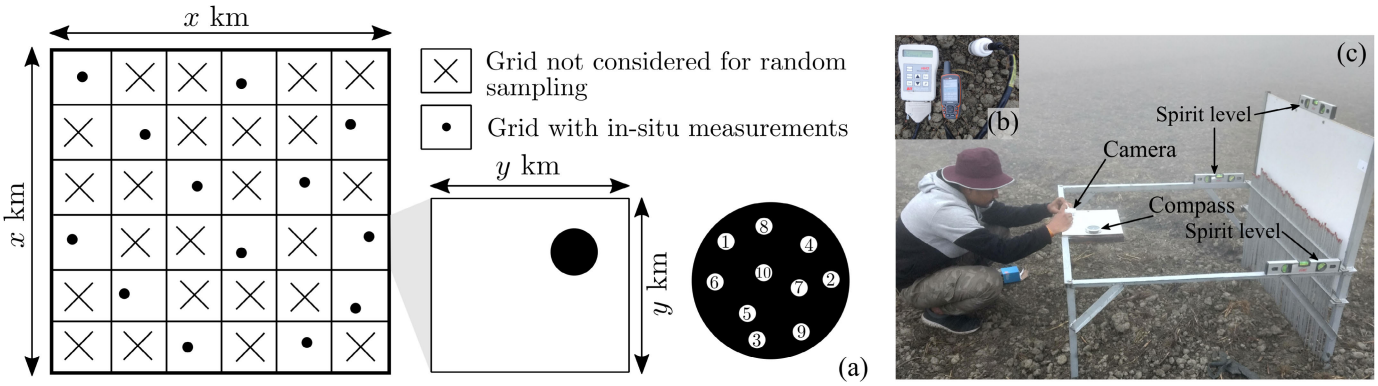

Fig. 3. (a) Schematic representation of the random sampling strategy of grid cells for the in situ measurements. The values of x and y for Kosi Fan are (24 km, 4 km). (b) Calibrated TDR for in situ moisture measurements and handheld GPS used in this study. (c) Mechanical pin-profilometer developed in-house for surface roughness measurements.

(also from Sentinel-1), and the NDVI derived from Sentinel-2. Elevation, longitude, and latitude information were obtained from the SRTM DEM. Given the high sensitivity of soil moisture detection to VH and VV polarizations, we generated two synthetic features for improved characterization; the VH/VV ratio and VH-VV difference using a linear data fusion approach. In total, nine features were derived from the satellite images and DEM data to serve as input variables for model training, along with the in situ measurements.

We conducted a field campaign in December (from 11th to 20th) 2019 on the Kosi Fan-in the Himalayan Foreland to measure the soil moisture and surface roughness. We used a calibrated time-domain reflectometry (TDR) and a mechanical pin-profilometer to measure surface soil moisture and roughness at 78 locations using a random grid sampling approach as illustrated in Fig. 3. The specifics of the study area, field measurements, and in situ data can be found in our previous works [48], [25], [37].

III. Model Development and Optimization

The initial step in developing a PIML model is the selection of a network architecture. There are a number of different network architectures that can be used for PIML, including convolutional neural networks, recurrent neural networks, and FFNNs. The choice of a network architecture will depend on the specific problem that is being solved. For instance, if the problem involves image processing, then a convolutional neural network may be a good choice. If a problem involves sequential data, then a recurrent neural network may be a good choice. On the other hand, if the problem involves nonsequential data, then an FFNN may be the best choice. In this study, we develop a learning bias PIML by incorporating the physics in the loss function of a 9 (inputs):3 (one hidden layer with three neurons):1 (output) fully connected FFNNs (Fig. 4). The input layer in this architecture is followed by a linear activation function (i.e., purelin), whereas the hidden and the output layer follow a logistic sigmoid activation function (i.e., logsig). The modification in the model involves matching the theoretically computed backscatter (VV_{I2EM}) with the satellite-derived backscatter (VV) along with the conventional term (1)

$$\text{Loss function} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(SM_{observed} - SM_{Predicted})}_{\text{Conventional term}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}(VV_{I2EM} - VV)}_{\text{Physics term}} \quad (1)$$

where n is the number of observations, $SM_{observed}$ is the in situ soil moisture, and $SM_{Predicted}$ is the predicted soil moisture. The primary motivation behind incorporating VV into the loss function is its high sensitivity to surface soil moisture, as demonstrated in previous studies [37], [49], [50]. It is important to note that the conventional component of the loss function provides robustness against uncertainties in the in situ observations by directly including the $SM_{observed}$ term. On the other hand, the modified component (physics term) is designed to counteract any uncertainties (noise) present in satellite images, specifically VV. The presence of uncertainties, whether in the in situ measurements or satellite images, can significantly impact the model's performance and hinder the convergence of the optimization process.

The time complexity of PIML is influenced by various factors, including the size of the dataset, the physics model's complexity, and the neural network architecture selection. As the size of the dataset increases, the neural network requires more training data, which in turn prolongs the model-tuning process. In addition, when the physics model is more intricate, it becomes more challenging for the neural network to learn, resulting in longer training times. The choice of neural network architecture also plays a role in the time complexity of PIML. More sophisticated architectures may necessitate extended training durations, but they have the potential to achieve higher levels of accuracy.

To simulate the co-polarized radar backscatter values (VV_{I2EM}), we used the improved integral equation model (I2EM) [51], which relates the backscattering coefficient to target and sensor properties such as surface roughness, soil dielectric, incidence angle, and wavelength [52], [53]. In this study, we focus solely on the co-polarized backscattering

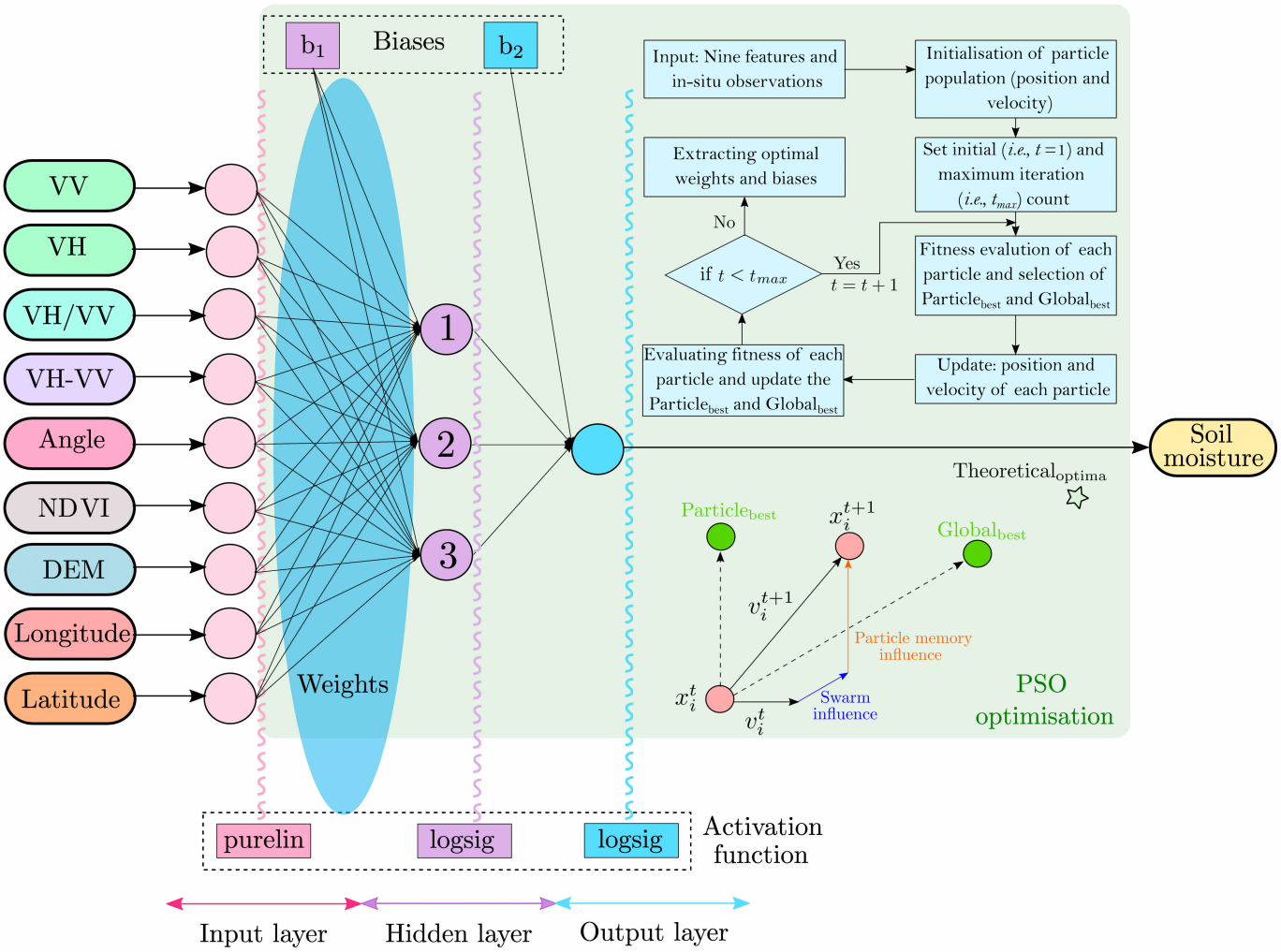

Fig. 4. Learning bias PIML architecture consists of nine inputs, a single hidden layer with three neurons. The model parameters are optimized using the PSO algorithm.

coefficient ($\mathrm{VV_{I2EM}}$) according to

$$\mathrm{VV_{I2EM}} = \frac{k^2}{4\pi} e^{-2k^2 s^2 \cos^2\theta} \sum_{n=1}^{+\infty} \left| I_{\mathrm{vv}}^n \right|^2 \frac{W_{\mathrm{ft}}^{(n)}(2k\sin\theta, 0)}{n!} \tag{2}$$

where k represents the wavenumber ($k = (2\pi/\lambda)$), λ is the wavelength, s is the surface roughness (rms height), θ is the incidence angle, and $W_{\mathrm{ft}}^{(n)}$ is the Fourier transform of surface autocorrelation function $\zeta(x, y)$. I_{vv}^n is given as

$$I_{\mathrm{vv}}^n = (2ks\cos\theta) f_{\mathrm{vv}} e^{(-k^2 s^2 \cos^\theta)} + (ks\cos\theta)^n F_{\mathrm{vv}} \tag{3}$$

where f_{vv} is given by

$$f_{\mathrm{vv}} = \frac{2\Psi_v}{\cos\theta} \tag{4}$$

where the vertically polarized Fresnel reflection coefficient is given by

$$\Psi_v = \frac{\epsilon\cos\theta - \sqrt{\epsilon - \sin^2\theta}}{\epsilon\cos\theta + \sqrt{\epsilon - \sin^2\theta}} \tag{5}$$

where ϵ represents the ground dielectric constant. F_{vv} and $W_{\mathrm{ft}}^{(n)}$ are given by

$$F_{\mathrm{vv}} = \frac{2\sin^2\theta}{\cos\theta}\left[\left(1 - \frac{\epsilon\cos^2\theta}{\epsilon - \sin^2\theta}\right)(1 - \Psi_v)^2 + \left(1 - \frac{1}{\epsilon}\right)(1 + \Psi_v)^2\right] \tag{6}$$

$$W_{\mathrm{ft}}^{(n)}(a, b) = \frac{1}{2\pi}\int\int \zeta^n(x, y) e^{-i(by+ax)} dxdy \tag{7}$$

where $\zeta(x, y)$ can be either exponential [(8)] or Gaussian [(9)] depending upon the roughness magnitude

$$\zeta(x) = e^{-\left(\frac{|x|}{\eta}\right)} \tag{8}$$

$$\zeta(x) = e^{-\left(\frac{x}{\eta}\right)^2} \tag{9}$$

where η is the correlation length. The I2EM model applies to a wide range of surface roughness from smooth to rough. The validity range is given by

$$ks \leq 3 \tag{10}$$

$$\left(\frac{(ks\cos\theta)^2}{\sqrt{0.46\ ks\eta}} \exp\left(-\sqrt{0.92\ ks\eta(1 - \sin\theta)}\right)\right) \ll 1. \tag{11}$$

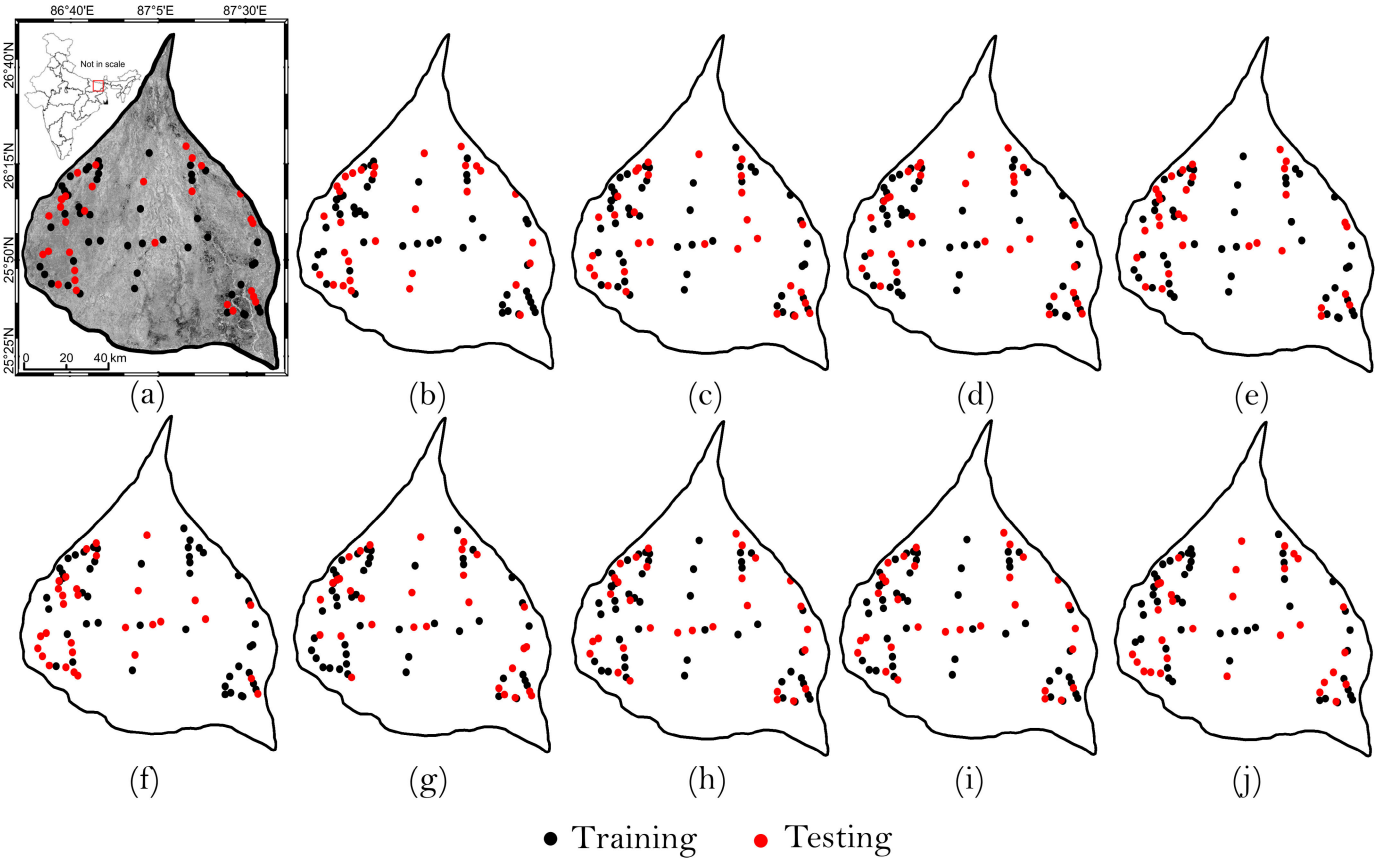

Fig. 5. Spatial distribution analysis using various sets of training and testing data on the Kosi Fan. For illustration, we have randomly displayed only ten (a)–(j) scenarios out of 30.

The in situ measured surface roughness parameters satisfy these constraints, thereby enabling the use of I2EM in the loss function.

Once we developed the learning bias PIML model, we used co- and cross-polarized backscatter images (VV and VH), incidence angle, VH/VV, and VH-VV from Sentinel-1 images, NDVI from Sentinel-2, latitude, longitude, and topographic information from the SRTM DEM as input features. To bring all the input features to a common level, we applied z-score scaling. Using 60% of the standardized input features and in situ measurements, we train the learning bias PIML model. During the training process, we optimize the values of model parameters (i.e., weights and biases) by minimizing (1) using particle swarm optimization (PSO) algorithm (swarm size = 9 and maximum iteration = 100). PSO, introduced by Kennedy and Eberhart [54], is a swarm intelligence-based optimization technique that offers several advantages, as it requires relatively few user-defined parameters compared to other optimization algorithms. The core optimization process relies on iterative updates governed by a set of equations, reducing computational complexity. It has demonstrated effectiveness in optimizing a wide range of theoretical and practical problems, as shown in previous studies [55], [56]. In PSO, a population of particles (or swarms), representing potential solutions, iteratively update their positions and velocities based on their own best solutions and the best solution found by the entire population in the following equations:

$$v_{\text{in}}^{t+1} = v_{\text{in}}^{t} + \chi_1 \cdot \rho_1 \cdot \left(\text{Particle}_{\text{best}}^{t} - x_{\text{in}}^{t}\right) + \chi_2 \cdot \rho_2 \cdot \left(\text{Global}_{\text{best}}^{t} - x_{\text{in}}^{t}\right) \quad (12)$$

$$x_{\text{in}}^{t+1} = x_{\text{in}}^{t} + v_{\text{in}}^{t+1}. \quad (13)$$

Each particle represents a possible solution, and it keeps track of two things: its own best solution found so far, called the "personal best" (represented by $\text{Particle}_{\text{best}}$) and the best solution found by the entire swarm, called the "global best" (represented by $\text{Global}_{\text{best}}$). The particles move around the search space, trying to improve their positions. Each particle considers two influences when deciding where to move next. First, its own experience and how close the particle is to its own best solution found so far. This is called the "cognitive component" (represented by χ_1). Second, the swarm's knowledge, that is, how close is the particle to the best solution found by the entire swarm so far? This is called the "social component" (represented by χ_2). To balance these influences, the particle uses random numbers between 0 and 1 (represented by ρ_1 and ρ_2). These random numbers add a bit of randomness to the movement, helping the particles explore new areas of the search space. Equations (12) and (13) describe how a particle's velocity and position are updated based on these influences. The same is illustrated in Fig. 4, where the particles move around the search space, constantly updating their position and

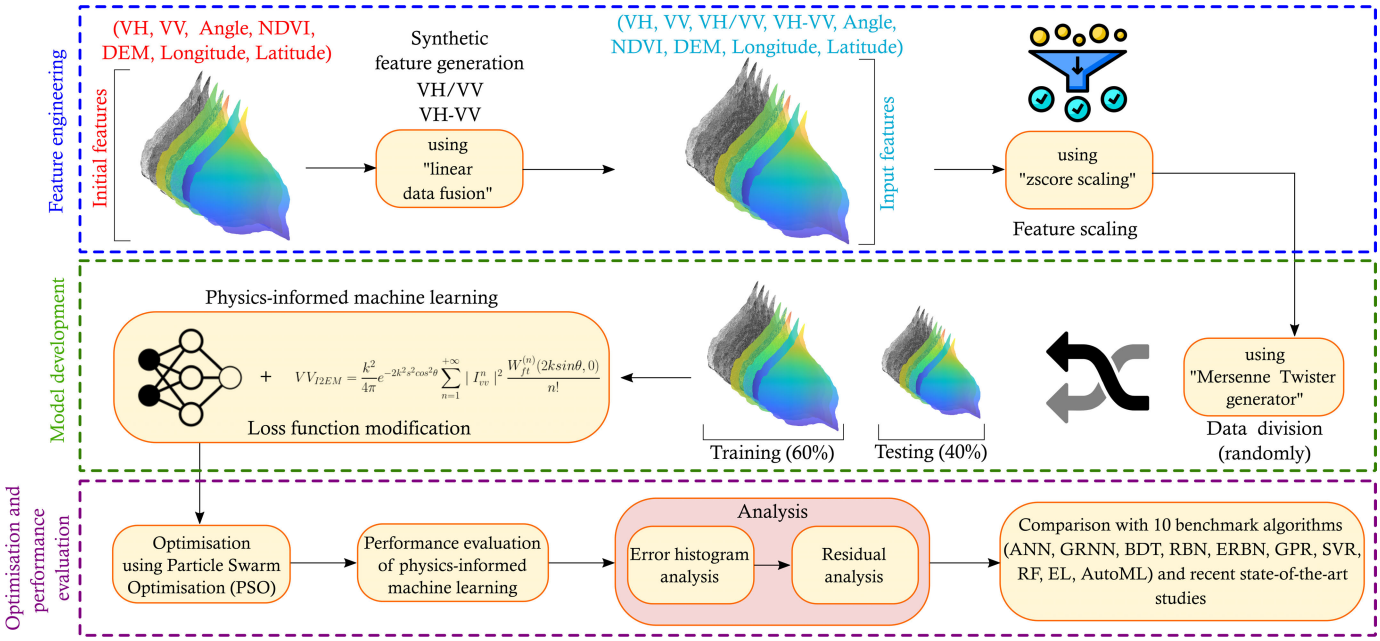

Fig. 6. Step-wise flowchart illustrating the soil moisture estimation through the PIML model.

speed. Two main factors influence this movement. The first is their own "memory," which pulls them closer to their personal best solution. The second is the "swarm's influence," which attracts them toward the best solution found by the whole group. Both these influences are visualized as lines parallel to the movement toward the personal and global best positions, respectively. Over time, the particles keep adjusting their position and speed based on these influences. This iterative process helps them get closer and closer to the ideal solution ($\text{Theoretical}_{\text{optima}}$) like a swarm gradually converging toward a food source. In a nutshell, the particles learn from their own experiences and the best solution found by the whole group. They use this knowledge to explore the search space and hopefully find the optimal solution to the problem.

Furthermore, assessing the uncertainty of the proposed model is crucial for making more informed decisions, as emphasized in previous studies [37], [57], [58], [59]. We performed an uncertainty quantification (spatial distribution analysis) by varying the input training data. We generated 30 independent training datasets by varying the random seed in the random generator and computed the corresponding variation in the model output (Fig. 5). For a comprehensive understanding of our methodology, we present a step-by-step flowchart in Fig. 6.

IV. Results and Discussion

A. Model Performance

To evaluate the prediction accuracy of the proposed learning bias PIML algorithm, we assessed its performance on an unseen dataset, which comprised 40% of the total data reserved for testing. A linear fit between the predicted and in situ observations was plotted, revealing that the predicted soil moisture closely aligns with the in situ values; R = 0.94, RMSE = 0.02 m^3/m^3, and bias = -0.03 m^3/m^3 [Fig. 7(a)]. All the observations are within the 95% confidence interval. Furthermore, we perform an error histogram analysis using ten bins to understand the model error distribution (i.e., error = predicted−observed). We plot the error involved during the training and testing phases in a stacked manner [Fig. 7(b)]. The vertical orange line denotes the zero-error line. Regions to the left and right of this line correspond to underestimation (negative errors) and overestimation (positive errors) zones, respectively. The resulting error histogram approximates a Gaussian distribution with a right-skew. The peak of the histogram is located near the zero-error line, indicating a strong model fit. The presence of positive errors represents overestimation. This is probably due to the numerous water-logging patches on the Kosi Fan [60]. This results high soil moisture levels on the Kosi Fan, especially following precipitation events, leading to rapid surface saturation. The model tends to overestimate soil moisture in these areas, likely due to this inherent variability in surface moisture retention and the region's fast response to rainfall [61]. In addition, incorporating radar physics in the model, specifically through Sentinel-1 backscatter signals, may contribute to the over-estimation. In wet surface conditions and areas with dense vegetation, which is common in the Kosi Fan, the backscatter signals can overrespond, capturing the waterlogged surface and vegetation moisture. This can lead to an amplified response in the predicted soil moisture values. We also plot the residuals between the fit and the in situ soil moisture [Fig. 7(c)]. Most residuals lie within the ±RMSE line and are stochastic, suggesting a good-fit model. Furthermore, we found an overall steady response for the uncertainty in the output corresponding to the variation in the input training datasets. We observed an average variation of 0.92 ± 0.01 in R, 0.03 ± 0.01 in RMSE, and -0.03 ± 0.01 in bias (Table II). Finally, we used the trained learning PIML to generate the spatial soil moisture map of the Kosi Fan (Fig. 8). It has a spatial resolution of

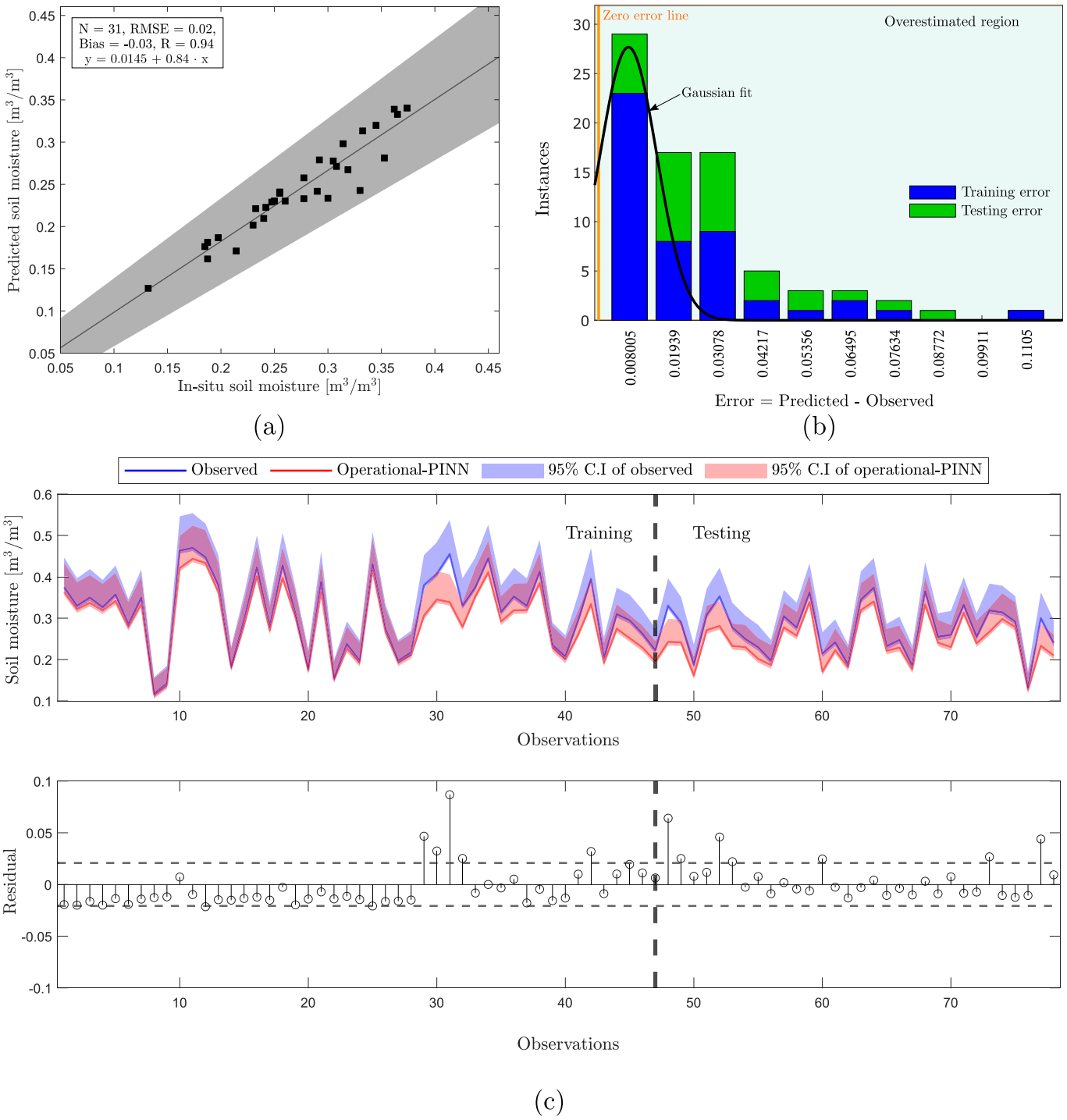

Fig. 7. (a) Linear regression plot showcasing the accuracy and correlation between predicted soil moisture and in situ measurements, with the gray shaded area representing the 95% confidence interval. (b) Error histogram analysis using ten bins. (c) Residual analysis, with the dashed line marking the testing ± RMSE.

60 m × 60 m. We observe high soil moisture near the western margin of the Kosi Fan, which is due to the topography of the region. The Kosi River flows near the western margin of the fan, contributing to the increased soil moisture. This area is also characterized by shallow groundwater, which explains the high moisture levels. The alluvial fans usually have a relatively high elevation at the fan axis and a lower elevation toward the margins. Drainage networks radiate from the central axis toward the fan margins, impacting soil moisture distribution. The proximity of the western margin to the Kosi River increases moisture content due to surface water and shallow groundwater interaction.

B. Benchmark Comparison

We evaluated the performance of the learning bias PIML algorithm against ten different machine learning models:

TABLE II
COMPARISON OF ACCURACY ACROSS VARIOUS SCENARIOS IN THE SPATIAL DISTRIBUTION ANALYSIS

Scenario	R	RMSE	Bias	Scenario	R	RMSE	Bias
Scenario 0	0.94	0.02	-0.03	Scenario 15	0.91	0.03	-0.03
Scenario 1	0.92	0.03	-0.03	Scenario 16	0.94	0.02	-0.04
Scenario 2	0.93	0.02	-0.02	Scenario 17	0.9	0.04	-0.04
Scenario 3	0.94	0.03	-0.02	Scenario 18	0.92	0.03	-0.03
Scenario 4	0.94	0.05	-0.04	Scenario 19	0.91	0.03	-0.02
Scenario 5	0.93	0.03	-0.03	Scenario 20	0.93	0.02	-0.03
Scenario 6	0.92	0.03	-0.02	Scenario 21	0.90	0.04	-0.03
Scenario 7	0.93	0.03	-0.03	Scenario 22	0.92	0.02	-0.02
Scenario 8	0.92	0.04	-0.03	Scenario 23	0.89	0.04	-0.04
Scenario 9	0.92	0.02	-0.06	Scenario 24	0.94	0.03	-0.03
Scenario 10	0.93	0.04	-0.04	Scenario 25	0.92	0.03	-0.02
Scenario 11	0.94	0.03	-0.03	Scenario 26	0.91	0.02	-0.03
Scenario 12	0.92	0.04	-0.02	Scenario 27	0.93	0.02	-0.02
Scenario 13	0.9	0.04	-0.04	Scenario 28	0.93	0.02	-0.03
Scenario 14	0.89	0.04	-0.04	Scenario 29	0.91	0.03	-0.03
				$\mu \pm \sigma$	0.92 ± 0.01	0.03 ± 0.01	-0.03 ± 0.01

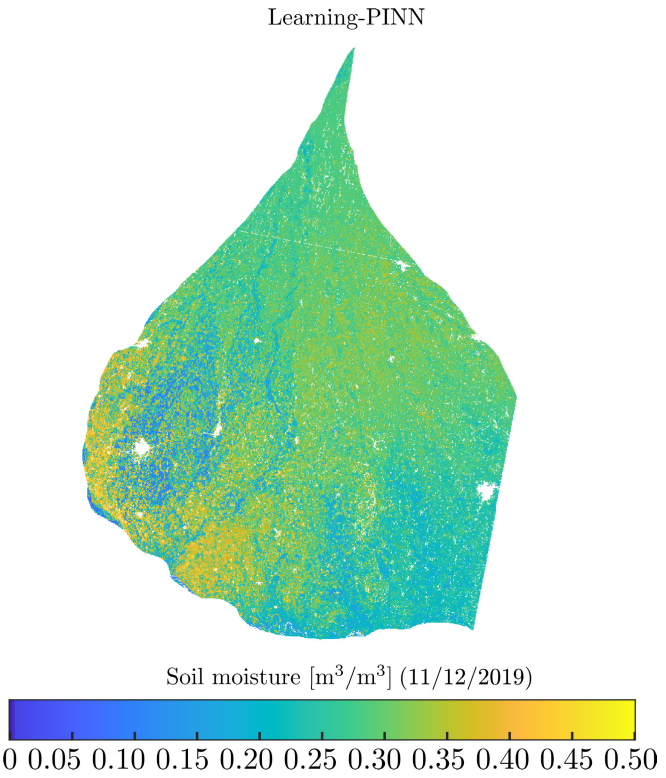

Fig. 8. Spatial soil moisture map derived from learning-PIML. The transparent pixel represents the invalid regions.

ANN, GRNN, binary decision tree (BDT), radial basis neural networks (RBNs), exact-RBN (ERBN), Gaussian process regression (GPR), SVR, RF, ensemble learning (EL) boosting, and AutoML. These benchmark algorithms are selected based on the bibliometric analysis, which selects the frequently used algorithms for soil moisture estimation. We use R, RMSE, bias, and refined index of agreement (d_r) [62] as the performance metrics to evaluate and compare the performance of the benchmark algorithms in the following equations:

$$R = \sqrt{1 - \frac{E_{\text{SSE}}}{E_{\text{SST}}}} \tag{14}$$

$$E_{\text{SSE}} = \sum (y_{\text{obs}} - y_{\text{pred}})^2 \tag{15}$$

$$E_{\text{SST}} = \sum (y_{\text{obs}} - \bar{y}_{\text{obs}})^2 \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_{\text{pred}} - y_{\text{obs}})^2} \tag{17}$$

$$\text{Bias} = \frac{\sum (y_{\text{pred}} - y_{\text{obs}})}{n} \tag{18}$$

$$d_r = \begin{cases} 1 - \dfrac{\sum_{i=1}^{n} |y_{\text{pred}} - y_{\text{obs}}|}{c \cdot \sum_{i=1}^{n} |y_{\text{obs}} - \overline{y_{\text{obs}}}|}, & \text{if } \omega \le c \cdot \sum_{i=1}^{n} |y_{\text{obs}} - \overline{y_{\text{obs}}}| \\ \dfrac{c \cdot \sum_{i=1}^{n} |y_{\text{obs}} - \overline{y_{\text{obs}}}|}{\sum_{i=1}^{n} |y_{\text{pred}} - y_{\text{obs}}|} - 1, & \text{if } \omega > c \cdot \sum_{i=1}^{n} |y_{\text{obs}} - \overline{y_{\text{obs}}}| \end{cases}$$

$$\omega = \sum_{i=1}^{n} |y_{\text{pred}} - y_{\text{obs}}| \tag{19}$$

where E_{SSE} and E_{SST} represent the sum of squares of errors and squares of the total, respectively. y_{pred} and y_{obs} represent the predicted and observed response variables, respectively. n is the total number of observations. Of these performance metrics, the R-value ranges from 0 to 1, with a higher value being preferable. RMSE ranges from 0 to ∞, where a lower value is considered better. The value bias ranges from $-\infty$ to $+\infty$, and a lower value is desirable. The value of d_r ranges from -1 to 1, and a higher value is desirable.

We apply these benchmark algorithms to the same datasets and evaluate their performance for comparison (Table III). Our analysis reveals that the learning bias PIML significantly outperforms all benchmark algorithms in terms of accuracy with R = 0.94, RMSE = 0.02, bias = -0.03, and $d_r = 0.86$.

C. Comparison With Previous Studies

We evaluate the accuracy of our proposed model by comparing it with leading studies that employ machine learning and remote sensing images for surface soil moisture prediction (Table IV). Most of these studies employed the RF and ANN to develop the predictive model [28], [29], [31], [32], [34], [63]. They use the in situ measured soil moisture in synergy with

TABLE III
COMPARISON WITH TEN BENCHMARK ALGORITHMS

Performance metrics	ANN	GRNN	BDT	RBN	ERBN	GPR	SVR	RF	EL (Boosting)	AutoML	(This study)
R	0.80	0.36	0.09	0.27	0.22	0.40	0.23	0.20	0.12	0.31	0.94
RMSE [m^3/m^3]	0.06	0.06	0.10	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.02
Bias [m^3/m^3]	-0.01	0.04	0.03	0.08	0.09	0.05	-0.08	0.04	0.07	0.02	-0.03
d_r	0.71	0.30	0.12	0.14	0.09	0.26	-0.02	0.22	-0.13	0.50	0.86

TABLE IV
COMPARISON WITH RECENT CUTTING-EDGE STUDIES IN SURFACE SOIL MOISTURE PREDICTION USING REMOTE SENSING AND MACHINE LEARNING TECHNIQUES

Publications	[36]	[35]	[34]	[33]	[32]	[31]	[30]	[29]	[63]	[28]	[25]	[37]	(This study)
Best performing algorithm	AutoML	SBC	ANN	GRNN	RF	RF	BRT	ANN	RF	RF	ANN	ANN	PIML
Performance metrics													
R	-	0.64	0.73	-	-	0.90	0.84	0.70 - 0.90	0.95	0.92	0.85	0.83	0.94
NS	>0.90	-	-	-	0.73	-	-	-	-	-	-	-	-
RMSE	0.04 cm^3/cm^3	0.075 m^3/m^3	0.035 cm^3/cm^3	0.069 cm^3/cm^3	-	0.049 cm^3/cm^3	-	3 - 7 %	1.48%	3%	0.05 m^3/m^3	0.05 m^3/m^3	0.019 m^3/m^3
MAE	-	-	-	-	-	-	3.78%	-	-	-	-	-	-
Bias	-	-0.009 m^3/m^3	-0.024 cm^3/cm^3	0.003 cm^3/cm^3	-	-	-	-	-	-	0 m^3/m^3	0 m^3/m^3	-0.03 m^3/m^3

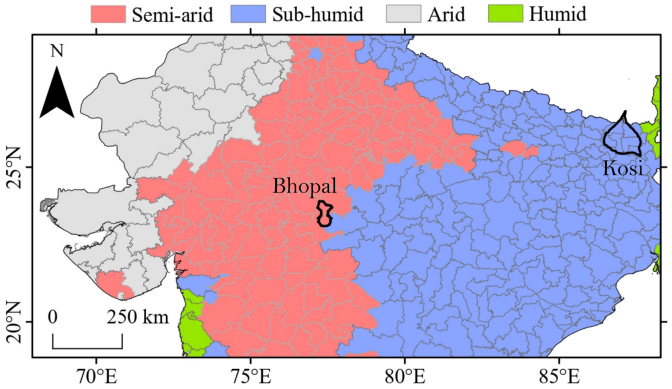

Fig. 9. Climatic zone classification of the Kosi and Bhopal study sites (Sources: ICAR-CRIDA).

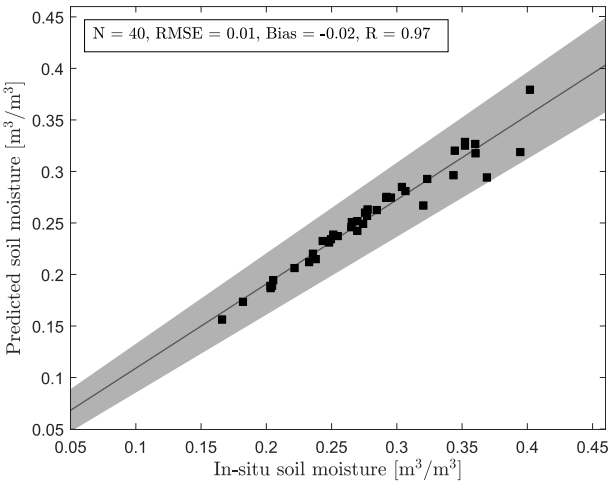

Fig. 10. Linear regression plot demonstrating the performance and alignment of the predicted soil moisture values with in situ measurements at the Bhopal site. The gray shaded region indicates the 95% confidence interval.

features extracted from the optical and microwave imagery to train the machine learning model. We found that the PIML outperforms all the recent studies that use satellite images as the input features. A combination of optical and microwave input features often results in the optimal result. Studies using UAV images show better accuracy than the satellite products [30], [63].

D. Performance Over Different Study Areas

We have tested the performance of the proposed PIML-SM model over a different study area with distinct climatic conditions to further evaluate its robustness. We selected Bhopal, a semi-arid region, which stands in contrast to Kosi, a humid subtropical area (Fig. 9). This allowed us to assess the model's generalizability across diverse climatic zones. Using in situ soil moisture observations from Bhopal region, as reported by Singh et al. [24], we trained the model and validated its performance on unseen testing datasets. The model demonstrated excellent accuracy, achieving the R-value of 0.97 and an RMSE of 0.01 m^3/m^3 (Fig. 10). These results highlight the model's ability to generalize effectively across regions with different climatic conditions, further strengthening its applicability for soil moisture estimation in diverse environments.

E. Model Interpretability, and Comparison With Black-Box Models

SAR backscatters exhibit high sensitivity to surface soil attributes, particularly soil moisture and roughness. This sensitivity varies depending on the polarization used. Due to the high sensitivity of VV toward soil moisture when compared to other available polarization, we consider the physics of the VV backscattering in the loss function to develop the PIML [64]. Moreover, VV is more sensitive to estimating surface soil moisture at bare soil as compared to VH [65]. In contrast, HH and HV are more sensitive to the surface roughness [66]. However, the sensitivity of the radar backscatter toward the soil attributes decreases with increasing vegetation. To address this limitation and incorporate the impact of vegetation, we introduce the NDVI as an input feature. NDVI allows the model to account for varying vegetation levels. Furthermore, incorporating the theoretically derived VV backscatter coefficient (VV_{I2EM}) alongside the satellite-derived VV measurement provides an additional layer of resilience against noise inherent in satellite imagery, such

as speckle noise. While preprocessing can reduce speckle, it cannot eliminate it [67]. Therefore, incorporating the satellite VV component acts as a shield against this noise.

As demonstrated by Singh and Gaurav [37], training a deep learning model often requires a substantial amount of data. Their fully connected feed-forward neural network, utilizing the same nine input features and 224 in situ observations (collected during field campaigns in 2019 and 2022) as the response variable for the Kosi Fan, achieved a testing accuracy of R = 0.80 and RMSE = 0.04 m^3/m^3. In contrast, this study proposes a PIML model that effectively leverages physical knowledge to learn from smaller datasets (trained on 78 in situ sites). This enables the PIML model to achieve competitive performance in soil moisture prediction (R = 0.94 and RMSE = 0.02). This advantage is particularly valuable when ground-truth data on soil moisture is limited or expensive to acquire. Furthermore, black-box models lack generalizability when applied to regional or global scales, particularly when encountering data outside their training range. In such scenarios with varying weather conditions or soil types, PIML models excel due to their reliance on established physical principles that hold across diverse contexts. This is crucial for robust predictions. In addition, black-box models can become entangled in spurious correlations, potentially leading to physically inconsistent predictions. For instance, a black-box model might correlate a specific Sentinel-1 band with high soil moisture. However, incorporating physics might reveal that this band is sensitive to surface roughness, a factor that can be high in both dry and wet conditions. This knowledge, when integrated into the model, can refine predictions significantly.

In conclusion, incorporating physical knowledge into machine learning models for soil moisture prediction fosters interpretable, generalizable, and physically consistent results, particularly when data scarcity is a challenge.

V. Conclusion

This study demonstrates that the data-driven models empowered by underlying physical principles can significantly enhance soil moisture prediction accuracy. We propose a learning bias PIML model for surface soil moisture prediction on the Kosi Fan. By incorporating a physics-aware loss function, the model achieves a 50% reduction in prediction error compared to the black-box model. Notably, this is accomplished with a 65% reduction in required data, highlighting the efficiency of the PIML approach. This translates to accurate soil moisture estimations from optical and microwave satellite images, achieving a substantial improvement in correlation (15%) compared to black-box models.

This study marks a step forward in improving the accuracy of soil moisture prediction with a limited number of training samples. The spatial distribution analysis of our model reveals its robustness in predicting soil moisture on the Kosi Fan. To further test its generalizability, we applied the model to the Bhopal site, which has a different climatic condition, and achieved acceptable accuracy, demonstrating the model's potential adaptability. While the PIML model may not be directly transferable to other regions without adjustment, these results suggest that the model can be effectively adapted to different geographical areas through transfer learning techniques, using the current model as a source domain and its optimized parameters as initial values for new regions.

Moving forward, this methodology could be applied to high-resolution images captured by UAVs. This would allow for monitoring soil moisture variability across both space and time. Such insights would be invaluable for agricultural productivity management, drought monitoring, and various other applications in Earth sciences.

Acknowledgment

The authors sincerely thank the Geohydrology Observatory facility at IISER Bhopal for providing the resources and environment that enabled them to carry out the initial experiments, which played a crucial role in shaping this study. They also deeply appreciate the editor and the anonymous reviewers for their valuable feedback and constructive suggestions, which have greatly enhanced the quality of this work.

CRediT Author Statement

Abhilash Singh led the conceptualization, methodology, data handling, and software development, while also drafting, visualizing, and refining the manuscript. **Kumar Gaurav** drove the methodology, software, and data management and provided critical insights, supervision, and project oversight.

Software and Data Availability

1) *Software Name:* Physics-informed machine learning for soil moisture estimation (PIML-SM).
2) *Contact Information:* abhilash.singh@ieee.org.
3) *Program Language:* MATLAB.
4) *Software Availability:* https://abhilashsingh.net/codes.html.
5) *Data Availability:*
   a) *Satellite Data:* https://scihub.copernicus.eu/.
   b) *In Situ Measurements:* In situ measurements have also been made publicly available and can be downloaded from our previous publications.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] S. A. Kannenberg, W. R. L. Anderegg, M. L. Barnes, M. P. Dannenberg, and A. K. Knapp, "Dominant role of soil moisture in mediating carbon and water fluxes in dryland ecosystems," *Nature Geosci.*, vol. 17, no. 1, pp. 38–43, Jan. 2024.

[2] R. Orth, "Global soil moisture data derived through machine learning trained with in-situ measurements," *Sci. Data*, vol. 8, no. 1, pp. 1–14, Jul. 2021.

[3] A. Singh, K. Gaurav, G. K. Sonkar, and C.-C. Lee, "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*, vol. 11, pp. 13605–13635, 2023.

[4] E. Njoku and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *J. Hydrol.*, vol. 184, nos. 1–2, pp. 101–129, Oct. 1996.

[5] W. Wagner et al., "Operational readiness of microwave remote sensing of soil moisture for hydrologic applications," *Hydrol. Res.*, vol. 38, no. 1, pp. 1–20, Feb. 2007.

[6] A. Singh, G. K. Meena, S. Kumar, and K. Gaurav, "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. Spatial Inf. Sci.*, vol. 4, no. 5, pp. 197–202, 2018.

[7] B. Li and S. P. Good, "Information-based uncertainty decomposition in dual-channel microwave remote sensing of soil moisture," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 9, pp. 5029–5045, Sep. 2021.

[8] I. P. Senanayake, I.-Y. Yeo, J. P. Walker, and G. R. Willgoose, "Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning," *Sci. Total Environ.*, vol. 776, Jul. 2021, Art. no. 145924.

[9] C. Ma et al., "Quantifying uncertainties in passive microwave remote sensing of soil moisture via a Bayesian probabilistic inversion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4406918.

[10] A. Singh, M. Niranjannaik, S. Kumar, and K. Gaurav, "Comparison of different dielectric models to estimate penetration depth of L- and S-band SAR signals into the ground surface," *Geographies*, vol. 2, no. 4, pp. 734–742, Nov. 2022.

[11] E. P. W. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Sci.*, vol. 13, no. 2, pp. 357–364, Mar. 1978.

[12] Y. Oh, K. Sarabandi, and F. T. Ulaby, "An empirical model and an inversion technique for radar scattering from bare soil surfaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 370–381, Mar. 1992.

[13] Y. Oh, K. Sarabandi, and F. T. Ulaby, "An inversion algorithm for retrieving soil moisture and surface roughness from polarimetric radar observation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, Apr. 1994, pp. 1582–1584.

[14] P. C. Dubois, J. van Zyl, and T. Engman, "Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 915–926, Jul. 1995.

[15] J. P. Walker and P. A. Troch, "Profile soil moisture estimation using the modified IEM," in *Proc. IGARSS*, vol. 3, May 1997, pp. 1263–1265.

[16] Y. Oh, K. Sarabandi, and F. T. Ulaby, "Semi-empirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1348–1355, Jun. 2002.

[17] Y. Oh, "Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 596–601, Mar. 2004.

[18] G. C. Topp, J. L. Davis, and A. P. Annan, "Electromagnetic determination of soil water content: Measurements in coaxial transmission lines," *Water Resour. Res.*, vol. 16, no. 3, pp. 574–582, Jun. 1980.

[19] M. R. Sahebi and J. Angles, "An inversion method based on multi-angular approaches for estimating bare soil surface parameters from RADARSAT-1," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 11, pp. 2355–2366, Nov. 2010.

[20] S. S. Rao et al., "Modified Dubois model for estimating soil moisture with dual polarized SAR data," *J. Indian Soc. Remote Sens.*, vol. 41, no. 4, pp. 865–872, Dec. 2013.

[21] X. Shen, K. Mao, Q. Qin, Y. Hong, and G. Zhang, "Bare surface soil moisture estimation using double-angle and dual-polarization L-band radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3931–3942, Jul. 2013.

[22] P. Thanabalan and R. Vidhya, "A synergistic approach for soil moisture estimation using modified Dubois model with dual-polarized SAR and optical satellite data," *Proc. SPIE*, vol. 9877, May 2016, Art. no. 98770Y.

[23] H. R. Mirsoleimani, M. R. Sahebi, N. Baghdadi, and M. E. Hajj, "Bare soil surface moisture retrieval from Sentinel-1 SAR data based on the calibrated IEM and Dubois models using neural networks," *Sensors*, vol. 19, no. 14, p. 3209, Jul. 2019.

[24] A. Singh, K. Gaurav, G. K. Meena, and S. Kumar, "Estimation of soil moisture applying modified Dubois model to sentinel-1; a regional study from central India," *Remote Sens.*, vol. 12, no. 14, p. 2266, Jul. 2020.

[25] A. Singh, M. N. Naik, and K. Gaurav, "Drainage congestion due to road network on the Kosi alluvial Fan, Himalayan Foreland," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102892.

[26] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, no. 1, pp. 69–80, Jan. 2010.

[27] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data," *Remote Sens.*, vol. 7, no. 12, pp. 16398–16421, Dec. 2015.

[28] I. Hajdu, I. Yule, and M. H. Dehghan-Shear, "Modelling of near-surface soil moisture using machine learning and multi-temporal Sentinel 1 images in New Zealand," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 1422–1425.

[29] E. Santi, M. Dabboor, S. Pettinato, and S. Paloscia, "Combining machine learning and compact polarimetry for estimating soil moisture from C-band SAR data," *Remote Sens.*, vol. 11, no. 20, p. 2451, Oct. 2019.

[30] S. N. Araya, A. Fryjoff-Hung, A. Anderson, J. H. Viers, and T. A. Ghezzehei, "Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 5, pp. 2739–2758, May 2021.

[31] V. Senyurek, F. Lei, D. Boyd, M. Kurum, A. C. Gurbuz, and R. Moorhead, "Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS," *Remote Sens.*, vol. 12, no. 7, p. 1168, Apr. 2020.

[32] H. Adab, R. Morbidelli, C. Saltalippi, M. Moradian, and G. A. F. Ghalhari, "Machine learning to estimate surface soil moisture from remote sensing data," *Water*, vol. 12, no. 11, p. 3223, Nov. 2020.

[33] Y. Cui et al., "A soil moisture spatial and temporal resolution improving algorithm based on multi-source remote sensing data and GRNN model," *Remote Sens.*, vol. 12, no. 3, p. 455, Feb. 2020.

[34] G. Ayehu, T. Tadesse, B. Gessesse, Y. Yigrem, and A. M. Melesse, "Combined use of Sentinel-1 SAR and Landsat sensors products for residual soil moisture retrieval over agricultural fields in the upper Blue Nile basin, Ethiopia," *Sensors*, vol. 20, no. 11, p. 3282, Jun. 2020.

[35] S. K. Chaudhary et al., "Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation," *Adv. Space Res.*, vol. 69, no. 4, pp. 1799–1812, Feb. 2022.

[36] E. Babaeian, S. Paheding, N. Siddique, V. K. Devabhaktuni, and M. Tuller, "Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning," *Remote Sens. Environ.*, vol. 260, Jul. 2021, Art. no. 112434.

[37] A. Singh and K. Gaurav, "Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images," *Sci. Rep.*, vol. 13, no. 1, p. 2251, Feb. 2023.

[38] A. Singh, M. Mehra, A. Kumar, M. Niranjannaik, D. Priya, and K. Gaurav, "Leveraging hybrid machine learning and data fusion for accurate mapping of malaria cases using meteorological variables in western India," *Intell. Syst. with Appl.*, vol. 17, Feb. 2023, Art. no. 200164.

[39] A. Kumar, K. Gaurav, A. Singh, and Z. M. Yaseen, "Assessment of machine learning models to predict daily streamflow in a semiarid river catchment," *Neural Comput. Appl.*, vol. 36, no. 21, pp. 13087–13106, Jul. 2024.

[40] A. Karpatne et al., "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2318–2331, Oct. 2017.

[41] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, May 2021.

[42] A. Singh, J. Amutha, J. Nagar, S. Sharma, and C.-C. Lee, "AutoML-ID: Automated machine learning model for intrusion detection using wireless sensor network," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, May 2022.

[43] A. Singh, J. Amutha, J. Nagar, and S. Sharma, "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.*, vol. 211, Jan. 2023, Art. no. 118588.

[44] J. Nagar, S. K. Chaturvedi, S. Soh, and A. Singh, "A machine learning approach to predict the k-coverage probability of wireless multihop networks considering boundary and shadowing effects," *Expert Syst. Appl.*, vol. 226, Sep. 2023, Art. no. 120160.

[45] A. V. Zuben, L. E. Perotti, and F. A. C. Viana, "Anatomically-guided deep learning for left ventricle geometry generation with uncertainty quantification based on short-axis MR images," *Eng. Appl. Artif. Intell.*, vol. 121, May 2023, Art. no. 106012.

[46] W. Zhu, R. Liang, J. Yang, Y. Cao, G. Fu, and Y. Cao, "A sub-region unet for weak defects segmentation with global information and mask-aware loss," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106011.

[47] A. M. Roy and S. Guha, "A data-driven physics-constrained deep learning computational framework for solving von Mises plasticity," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106049.
[48] A. Singh, K. Gaurav, A. K. Rai, and Z. Beg, "Machine learning to estimate surface roughness from satellite images," *Remote Sens.*, vol. 13, no. 19, p. 3794, Sep. 2021.
[49] D. Palmisano, F. Mattia, A. Balenzano, G. Satalino, N. Pierdicca, and A. V. M. Guarnieri, "Sentinel-1 sensitivity to soil moisture at high incidence angle and the impact on retrieval over seasonal crops," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7308–7321, Sep. 2021.
[50] R. Meyer et al., "Exploring the combined use of SMAP and Sentinel-1 data for downscaling soil moisture beyond the 1 km scale," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 13, pp. 3337–3357, Jul. 2022.
[51] F. T. Ulaby et al., *Microwave Radar and Radiometric Remote Sensing*, vol. 4. Ann Arbor, MI, USA: University of Michigan Press, 2014.
[52] A. K. Fung, Z. Li, and K. S. Chen, "Backscattering from a randomly rough dielectric surface," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 356–369, Mar. 1992.
[53] N. Baghdadi et al., "A potential use for the C-band polarimetric SAR parameters to characterize the soil surface over bare agriculture fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3844–3858, Oct. 2012.
[54] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, vol. 4, Aug. 1995, pp. 1942–1948.
[55] A. Singh, S. Sharma, and J. Singh, "Nature-inspired algorithms for wireless sensor networks: A comprehensive survey," *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100342.
[56] X. Zhang, D. Zou, and X. Shen, "A novel simple particle swarm optimization algorithm for global optimization," *Mathematics*, vol. 6, no. 12, p. 287, Nov. 2018.
[57] J. Donnelly, S. Abolfathi, J. Pearson, O. Chatrabgoun, and A. Daneshkhah, "Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model," *Water Res.*, vol. 225, Oct. 2022, Art. no. 119100.
[58] B. Ghiasi et al., "Uncertainty quantification of granular computing-neural network model for prediction of pollutant longitudinal dispersion coefficient in aquatic streams," *Sci. Rep.*, vol. 12, no. 1, p. 4610, Mar. 2022.
[59] R. Noori et al., "An efficient data driven-based model for prediction of the total sediment load in rivers," *Hydrology*, vol. 9, no. 2, p. 36, Feb. 2022.
[60] M. Niranjannaik, K. Gaurav, A. Singh, and A. K. Singh, "Dynamics of waterlogging and drainage congestion on the Kosi Fan, Himalayan Foreland," *Remote Sens. Appl., Soc. Environ.*, vol. 33, Jan. 2024, Art. no. 101113.
[61] X. Ling et al., "Comprehensive evaluation of satellite-based and reanalysis soil moisture products using in situ observations over China," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 7, pp. 4209–4229, Jul. 2021.
[62] C. J. Willmott, S. M. Robeson, and K. Matsuura, "A refined index of model performance," *Int. J. Climatol.*, vol. 32, no. 13, pp. 2088–2094, Nov. 2012.
[63] X. Ge et al., "Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring," *PeerJ*, vol. 7, p. e6926, May 2019.
[64] M. Autret, R. Bernard, and D. Vidal-Madjar, "Theoretical study of the sensitivity of the microwave backscattering coefficient to the soil surface parameters," *Int. J. Remote Sens.*, vol. 10, no. 1, pp. 171–179, Jan. 1989.
[65] A. K. Hoskera, G. Nico, M. Irshad Ahmed, and A. Whitbread, "Accuracies of soil moisture estimations using a semi-empirical model over bare soil agricultural croplands from Sentinel-1 SAR data," *Remote Sens.*, vol. 12, no. 10, p. 1664, May 2020.
[66] N. Holah, N. Baghdadi, M. Zribi, A. Bruand, and C. King, "Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields," *Remote Sens. Environ.*, vol. 96, no. 1, pp. 78–86, May 2005.
[67] H. Choi and J. Jeong, "Speckle noise reduction technique for SAR images using statistical characteristics of speckle noise and discrete wavelet transform," *Remote Sens.*, vol. 11, no. 10, p. 1184, May 2019.

Abhilash Singh (Member, IEEE) received the Ph.D. degree from the Indian Institute of Science Education and Research, Bhopal, India, in 2024.

His current research interest includes coupling machine learning algorithms with satellite imagery to exploit satellite observations (soil moisture, surface roughness, and river discharge).

Dr. Singh is a member of EGU, AGU, ISPRS, and the Indian Radio Science Society (InRaSS). He was a recipient of the Vice-Chancellor Gold Medal Award from the University for being a first-rank holder in his bachelor's and master's. He received the prestigious "DST-INSPIRE" Fellowship to carry out his doctoral degree from the Department of Science and Technology (DST), India's Ministry of Science and Technology. He also received the DAAD Fellowship and several grants from the American Geophysical Union (AGU) and the European Geophysical Union (EGU).

Kumar Gaurav (Member, IEEE) received the M.Sc. degree in geoinformatics from the Indian Institute of Remote Sensing (IIRS), Dehradun, India, and the University of Twente (ITC), Enschede, The Netherlands, in 2009, and the Ph.D. degree from the Geological Fluid Dynamics Laboratory, Institut de Physique du Globe de Paris (IPGP), Paris, France, in 2016.

He is working as an Associate Professor with the Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research, Bhopal, India. He is a geomorphologist working on fluvial problems. He is interested in learning rivers' dynamics and morphology, drainage congestion's effect on flood inundation, and the application of remote sensing to monitor soil moisture and stream flow. He combines mathematical theory, physical reasoning, field observations, machine learning, and digital image analysis to recognize the necessary ingredients to model them.

Dr. Gaurav is also an Elected Member of the National Academy of Science, India (NASI) in Earth Sciences.