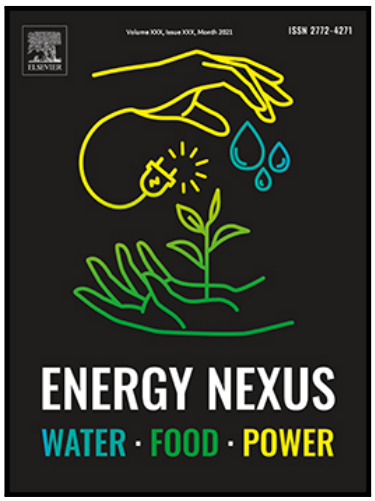
Journal Pre-proof



Predictive Modeling of soil salinity integrating remote sensing and soil variables: An ensembled deep learning approach

Sana Arshad , Jamil Hasan Kazmi , Endre Harsányi , Farheen Nazli , Waseem Hassan , Saima Shaikh , Main Al-Dalahme , Safwan Mohammed

PII: S2772-4271(25)00015-4
DOI: https://doi.org/10.1016/j.nexus.2025.100374
Reference: NEXUS 100374

To appear in: *Energy Nexus*

Received date: 3 September 2024
Revised date: 18 November 2024
Accepted date: 7 February 2025

Please cite this article as: Sana Arshad , Jamil Hasan Kazmi , Endre Harsányi , Farheen Nazli , Waseem Hassan , Saima Shaikh , Main Al-Dalahme , Safwan Mohammed , Predictive Modeling of soil salinity integrating remote sensing and soil variables: An ensembled deep learning approach, *Energy Nexus* (2025), doi: https://doi.org/10.1016/j.nexus.2025.100374

Predictive Modeling of soil salinity integrating remote sensing and soil variables: An ensembled deep learning approach

Sana Arshad[1], Jamil Hasan Kazmi[2], Endre Harsányi[3,4], Farheen Nazli[5], Waseem Hassan[6], Saima Shaikh[2], Main Al-Dalahme[7], Safwan Mohammed*[3,4]

[1]Department of Geography, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

[2]Department of Geography, University of Karachi, Karachi 75270, Pakistan

[3] Institute of Land Use, Technical and Precision Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary

[4] Institutes for Agricultural Research and Educational Farm, University of Debrecen, Böszörményi 138, 4032 Debrecen, Hungary

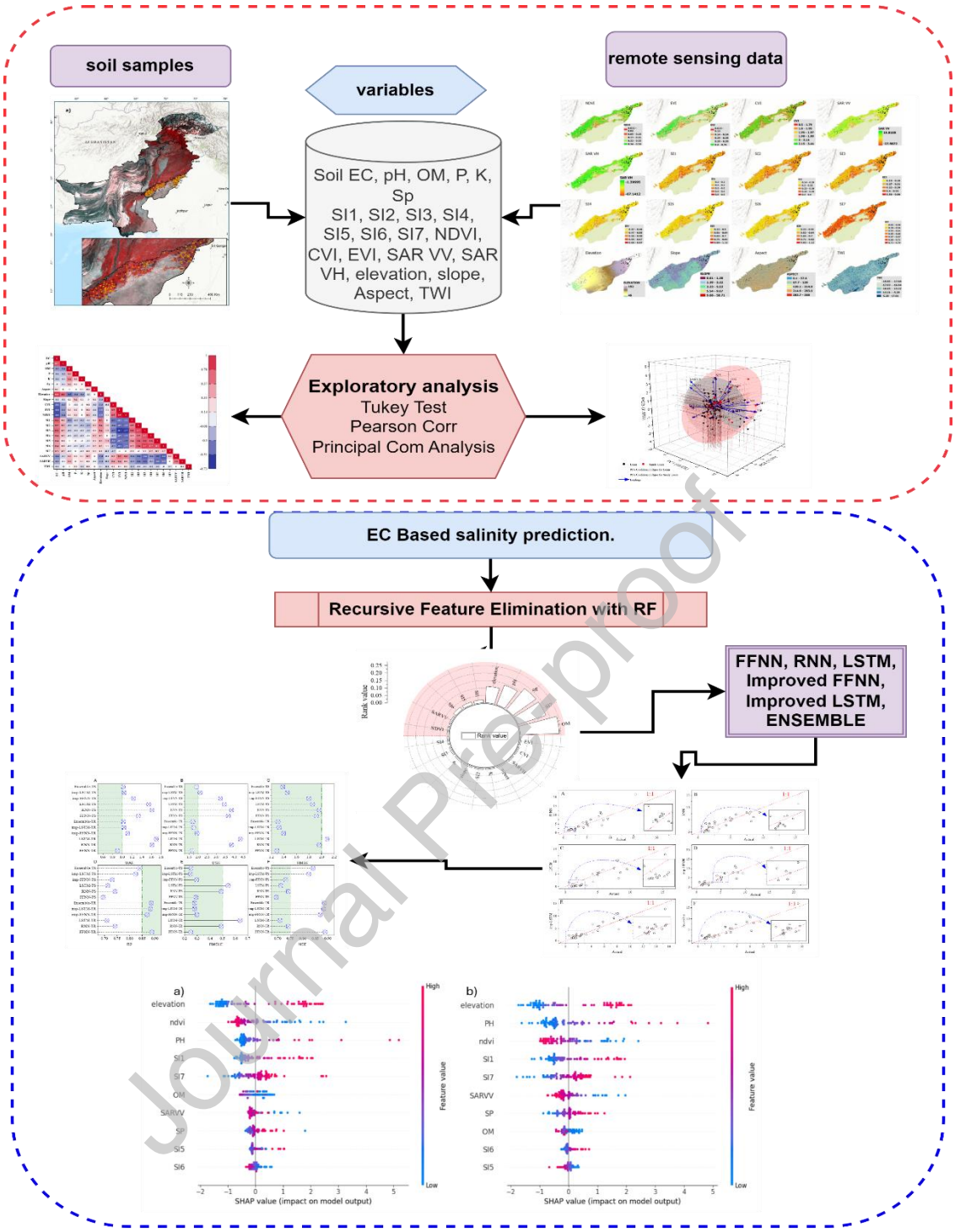
[5] Institute of Agro industry and Environment, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

[6] Soil and Water Testing Laboratory for Research Bahawalpur, Bahawalpur 63100, Pakistan

[7] Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan

(*safwan@agr.unideb.hu)

GRAPHICAL ABSTRACT



HIGHLIGHTS

1. Electrical Conductivity (EC) of the soil samples ranged between 0.57dS/m and 11.5 dS/m.
2. Soil EC was highly positively correlated with the pH, elevation, and salinity indices.
3. The ensemble of the improved FFNN and LSTM architectures with regularization outperformed (R^2 = 0.84, RMSE = 1.38, MAE = 1.01) in accurate salinity predictions.
4. SHAP revealed that elevation, pH, NDVI, SI-I, and CRSI had a high impact on salinity predictions.

Abstract

Accurate predictions of soil salinity can significantly contribute to achieving the UN-Sustainable Development Goal (SDG-2) of ensuring 'zero hunger.' From this perspective, the current research aimed to predict soil electrical conductivity (EC) from remote sensing and soil data using advanced deep learning (DL) architectures. A total of 109 soil samples were analyzed for agricultural land use in the Middle Indus Basin of Pakistan. Seven salinity indices (SI-1 to SI-7) were derived from the 10m to 20m wavelength bands of Sentinel-2, along with vegetation and topographic covariates. Initially, Recursive Feature Elimination was implemented as a feature-selection method to select the most effective predictors. Subsequently, deep learning architectures, including a Feedforward Neural Network (FFNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM), were employed to predict soil salinity. Research findings showed that EC ranged between 0.57dS/m to 11.5 dS/m in the study area. The evaluation metrics of the DL models revealed that a simple FFNN with three fully connected dense layers achieved the highest R^2 = 0.88 for model training. However, the ensemble of improved FFNN and LSTM outperformed with the highest R^2 and NSE = 0.84, and the lowest RMSE and MAE = 1.38 and 1.01, respectively, on the testing dataset. Optimized deep learning architectures with adjustments to the learning rate, dropout rate, and activation functions achieved the highest prediction accuracy with the lowest validation loss. Finally, SHapely Additive exPlanations (SHAP) revealed that elevation, pH, NDVI, SI-1, and SI-7 had highly significant impacts on EC predictions. This research provides insight into implementing advanced and interpretable DL architectures, supporting informed decision-making by agricultural stakeholders.

Key words: Electrical Conductivity, Canopy Response Salinity Index, Feed Forward Neural Network, Pakistan

Introduction

Soil salinity is a complex and dynamic phenomenon varying at different spatial and temporal scales [1]. Globally, more than 900 million hectares of agricultural land are subjected to different types of soil salinization (i.e., sodium (Na^+), potassium (K^+) sulfate (SO_4^{-2}), and chloride (Cl^-)) [2]. Soil salinity is linked to both primary (natural) and secondary (human-induced) factors. Natural factors, such as fluvial, aeolian, and stream flows, are mostly linked to rainfall deposition, geological structure, and geomorphic phenomenon [3]. However, the greatest cause of secondary soil salinization is related to intensive irrigation practices from the surface or groundwater. In extreme weather conditions, such as high temperatures, the soil experiences considerable evaporative water loss, leading to subsurface accumulation of evaporative salts in the soil. In addition to climatic factors, it is also dependent on vegetation type and abundance, land use and terrain, and the groundwater table in different geographical settings [4]. Soil salinization in irrigated farming regions is lowering the crop yield, putting pressure on sustainable food production and hindering the achievement of Sustainable Development Goal (SDG-2) of 'zero hunger.'[5, 6]. In arid and semi-arid regions of the world, crop yield losses due to soil salinity are estimated to range from 18 to 43 percent [7]. For instance, Eswar, Karuppusamy [8] reported that climate change is the main influencer of soil salinity for all background reasons. Increasing temperature and deficient rainfall exacerbate the pressure on available water resources for irrigation purposes which ultimately aligns with the salinization problem [9].

The electrical conductivity (EC) (dS/m) of soil is highly related to dissolved soil salts and is considered a rapid parameter for measuring soil salinity at local and regional levels

[10]. Previous research has explored various other soil-related factors, such as pH, organic matter, soil texture and structure, minerals, and saturation percentage, as the main determinants of EC-based soil salinity [11-13]. Conventionally, the spatial analysis of soil salinity was very limited, laborious, and time-consuming [14] until the development of geospatial tools and methods [15]. Moreover, laboratory-based measurements of soil salinity are site-specific and have a specific duration [16]. The progressive development of remote sensing methods for predicting and mapping soil salinity is indispensable in this regard [17, 18]. The visible and infrared bands of optical sensors and polarized backscattering of radar have been found to influence soil salinity analysis [19, 20]. Because of the high spatiotemporal variability and land use dynamics in saline areas, the choice of the spatial and spectral properties of satellites/sensors is crucial for soil salinity monitoring [21]. For instance, Avdan, Kaplan [22] outranked the performance of optical sentinel-2 and Landsat satellites over a high-resolution Planet Scope to avoid complexity in salinity predictions. Similarly, El Harti, Lhissou [23] developed a new OLI-salinity index from the Landsat satellite to estimate soil electrical conductivity. Furthermore, Scudiero, Skaggs [24] accurately predicted the linear relationship between field-based EC measurements and the landsat-derived canopy response salinity index (CRSI). Moreover, Ge, Ding [25] used hyperspectral imaging system for estimating soil salinity in China. Hence, a range of ratio- and intensity-based spectral indices are being utilized to map subsurface soil salinization and predict ground-based EC measurements [24, 26]. **Table 1** provides a comprehensive review of recent soil salinity prediction studies that have used ground-based and satellite measurements. In recent decades, the advent of artificial intelligence has enhanced the significance of salinity predictions using machine learning (ML) algorithms such as Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting Regression (GBR) [27-30].

Table 1. Recent review of soil salinity prediction studies integrating remote sensing and machine learning method

Objective	Region	Predictors	Method	Findings	Reference
Predicting soil salinity (EC) and sodicity (SAR)	Arid continental region in China	Sentinel-1 and 2, land use, topographic variables	Random Forest, SVM, and Cubist in four scenarios	Adding topographical variables with Sentinel-1 improved the prediction accuracy in RF	[31]
Predicting soil salinity	Temperate and flat	Remote sensing vegetation and	Decision Trees (DT),	DT outperformed	[32]

(EC) from remote sensing indices	topographic region of northern Iran	salinity indices, topographic, temperature and precipitation	Random Forest, Support Vector Regression, and XGBoost	with high prediction accuracy of soil salinity	
Predicting and mapping soil salinity (EC) from multisource data	Yellow river delta, China	Soil, remote sensing, meteorological, LULC, and sea data	Kriging regression, Random Forest, CatBoost	CatBoost showed high prediction accuracy, soil Na and NDVI proved significant predictors	[33]
Predicting soil salt content (SSC) defined by sum of 8 ionic species (h/kg)	Arid continental agriculture region of China	Soil information, topographical and surface, RS salinity and veg indices	Mix-up augmentation, Light GBM with Bayesian optimization	Augmentation improved the model performance solving overfitting, DEM is the most influential variable.	[34]
Predicting exchangeable sodium percent (ESP) and soil electrical conductivity (EC)	Irrigated agriculture land of Bolivia	Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, CO_3^{2-}, HCO_3^{2-}, SO_4^{2-},	PLS, RF, and SVM	RF outperformed followed by SVM.	[35]
Predicting soil salinity indices i.e., potential salinity (PS), sodium & magnesium adsorption ratio (SAR & MAR), ESP and TDI	Saline agricultural land of China.	Electrical conductivity, soil water and temperature, & potential hydrogen	SVM, RF, XGB,	XGB outperformed with EC, soil water content (SWC) and temperature (T) with highest R^2	[29]
Salinity prediction based on sampled EC measurement	Arid UAE	58 spectral indices were used from sentinel-2 satellite	M5P, Random Forest, Linear regression, instance base learning (IBk)	M5P and IBk outperformed for salinity prediction in SWIR2 region of sentinel-2 bands	[36]

Soil salinity prediction from sentinel-2 data	Lake Basin China	Topographical, soil, remote sensing indices, environmental, LULC	Boruta algorithm with machine learning classifiers and regressors	GBR outperformed for accurate prediction and B5, B6, TWI, DEM, temp and soil moisture identified as imp variables	[37]
Soil salinity mapping and prediction at various depths based on sampled EC	Oasis region Chine	Soil, remote sensing (Landsat) vegetation and soil indices, topographical (DEM) factors)	ATM, CART, SVR, GPR, SGT, LMSLR, M5, MARS, REPTress, MLP-ANN, RF	LMSLR and RF provided high accuracy and stability in soil salinity mapping and prediction,	[38]

Arid irrigated lands in Asia are among the most salt-affected regions in the world, with more than 200 Mha in Central Asia and 82 Mha in South Asia [8]. Saline soil in South Asia is mainly attributed to the high risk of soil erosion, declining water table, salted groundwater, and inappropriate irrigation practices [39].

Pakistan is one of the countries in South Asia that is most highly exposed to soil salinity problems in arid regions. It is estimated that approximately 21% of irrigated Indus plains are affected by different levels of salinity and sodicity, because irrigation practices are exaggerated in regions where groundwater is already saline [40]. The expansion of Indus basin irrigation in the country brought 30 million tons of salt through its canal commands to irrigated lands in previous years [41]. Overall, one-third of the agricultural land area deteriorates due to salinity problems, resulting in a 64% yield loss [42]. In this context, only a few studies have been conducted to address this issue in Pakistan, integrating remote sensing predictors to monitor soil salinization in machine learning environments [43, 44]. However, a rigorous review of the relevant literature also revealed a lack of detailed assessments of soil salinity in the middle Indus region of Pakistan, which is prone to arid and semi-arid conditions [45]. Moreover, from the perspective of machine learning applications, previous research has mainly focused on utilizing classical machine learning algorithms such as random forest, support vector regression, and boosting regressions (**Table 1**). The potential of deep learning algorithms for soil salinity prediction across different spatiotemporal scales needs to be explored. Moreover, a few previous studies (**Table 1**) utilized a combination of

high-resolution multispectral and Synthetic Aperture Radar (SAR)-based polarization properties to predict soil EC in semi-arid irrigated farming regions.

To address these gaps, our study provides a more detailed assessment of EC-based soil salinity prediction in the lower middle Indus region of Pakistan by integrating 16 high-resolution remote sensors and five field-collected soil indicators in the same time frame. The study also leveraged the use of advanced machine and deep learning algorithms, namely the Feedforward Neural Network (FFNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), improved FFNN, improved LSTM and Ensemble of FFNN and LSTM for accurate soil salinity predictions. Machine learning modeling is initiated by Recursive Feature Elimination (RFE) combined with a Random Forest algorithm for accurate feature selection. The modeling results were interpreted by SHapely Additive exPlanations (SHAP) analysis based on game theory, which provides a more comprehensive impact of individual features on soil salinity. In summary, this study aimed to 1) evaluate the soil characteristics and salinity in the lower middle Indus region of Pakistan, 2) investigate the performance of remote sensing indicators in predicting soil salinity, and 3) assess the performance of advanced machine and deep learning in predicting soil salinity using observed and remote sensing data.

2. Study area

The Central or Middle Indus Basin of Pakistan, known as the Sulaiman sub-basin, encompasses the Punjab Platform, the Sulaiman Depression, and the Sulaiman Fold Belt [46]. The soil parent material of the basin is inherently saline, with shallow groundwater depth. Generally, the soil of the upper Indus Basin is comprised of siliceous and argillaceous mudstone, whereas the middle and lower basins are rich in calcareous rocks [47]. Currently, the collected soil samples belong to the Punjab platform of the lower middle Indus Basin of Pakistan (27°N–29.7°N, 69.6°E – 73.8°E), which mainly comprises three districts: Bahawalnagar, Bahawalpur, and Rahim Yar Khan (**Fig 1a**). The elevation of the region ranges from 195 m in the northeast to less than 50 m in the southwest (**Fig 1b**). The major land cover of the region comprises agricultural land, barren and built-up land, and temporary water bodies in the form of canals and dry beds in the Sutlej River (**Fig 1c**). The climate of the region is hot and dry, with a mean annual rainfall of less than 200 mm, resulting in arid to semiarid conditions. agricultural pattern of the region is seasonal, with wheat as winter-dominant and cotton, sugarcane, and rice as summer-dominant crops [48]. The parent material of the region is composed of mixed calcareous alluvium with sand dunes. The

dominant soil texture is loam, sandy loam, and clay loam, belonging to Calcaric Rhegosols, Haplic Yermosols, and Calcaric Fluvisols, respectively, of the FAO World Agricultural Soil Classification [49]. Croplands are mainly dependent on the Indus Basin canal irrigation system, which is the primary cause of secondary salinization. It is estimated that the Indus River, along with its tributaries, deposits 31 million tons of salt annually in its basin [50]. A recent evaluation of the soil profile in the region revealed that 28% of the middle to lower Indus Basin of the country is saline and 20% is sodic, increasing the vulnerability of agricultural food production systems [51].

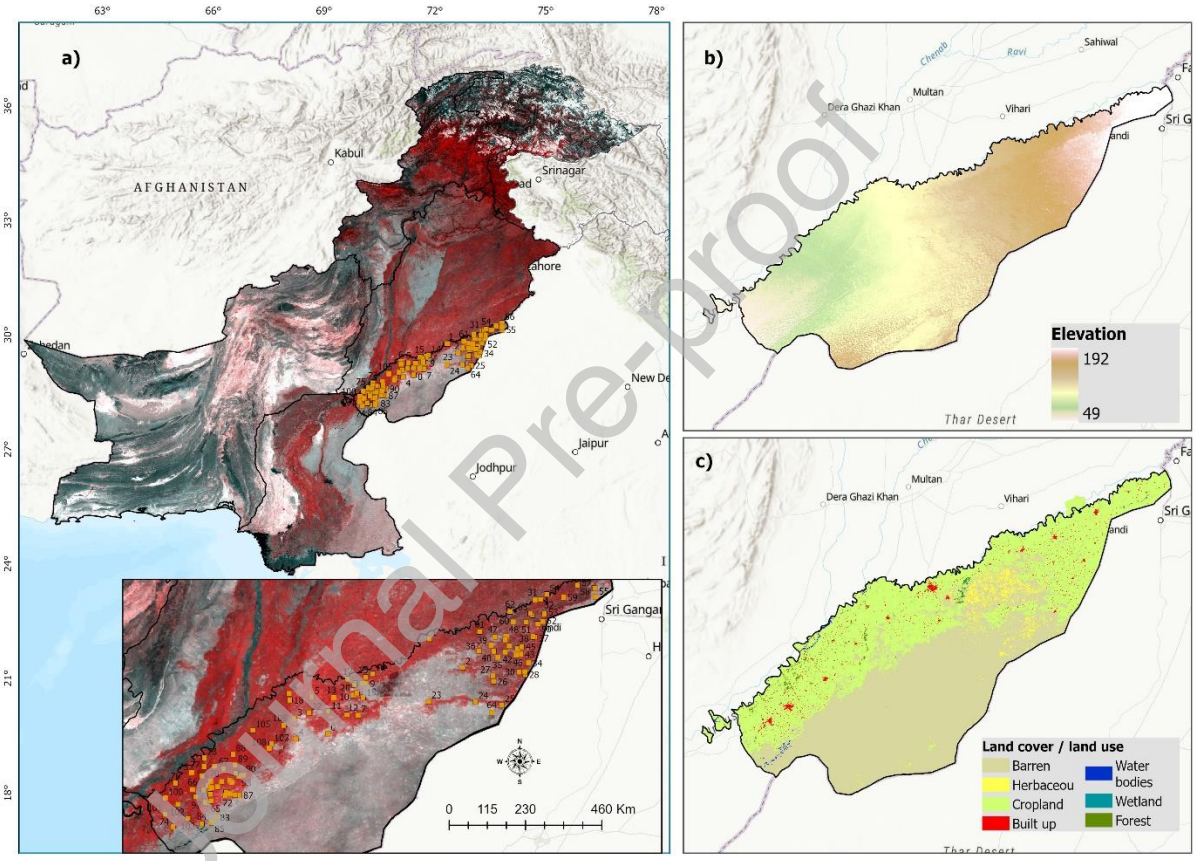


Fig 1 a) Location of study area in Pakistan with soil sample distribution presented on False Color Composite of Sentinel 2B. **b)** elevation in meters derived from: USGS/SRTMGL1-003. **c)** Land use land cover derived from: COPERNICUS/Landcover/100m/Proba-V-C3/Global

3. Material and Method

3.1. Data collection

3.1.1. Soil sampling

Soil samples were collected at various intervals throughout 2022 and aligned with seasonal crop rotations. A total of 109 evenly distributed samples were collected from agricultural

fields, with a focus on capturing diverse soil properties across the study area (**Fig 1a**). Samples were collected from the top 0–15 cm of the soil to ensure a good representation of the agricultural practices and patterns of crop rotation. To avoid redundancy, no samples were collected from agricultural fields smaller than 10 acres in size. The selection of sample sites also focused on accessibility and coverage of dominant croplands across the study area (**Fig 1a**). This approach provides a good representation of the soil conditions in agricultural land use for predicting soil salinity. The geographic coordinates of the sample locations were obtained using a manually handled GPS device (Garmin GPS 60) with a positional accuracy of less than 5 m. All samples were collected in labelled polythene bags and dried at 30–35 °C at room temperature with a maximum humidity of less than 50%, ground, and sieved through a 2 mm sifter. Subsequently, a 1:5 soil–water suspension was prepared using deionized water. The electrical conductivity of the soil represented the concentration of soluble slats in the suspension, which was measured using a digital Jenway conductivity meter. The pH was measured using a pH meter. Soil texture, organic matter (OM), available phosphorus (P), and extractable potassium (K) were determined following the methodology described by Ali, Shabaan [52]. Finally, all soil attributes EC (dS/m), pH, OM (%), P (ppm), K (ppm), and saturation (%) (**Table 3**) were mapped and classified in the study region by employing Inverse Distance Weighting IDW spatial interpolation in ArcGIS (**Fig 2A**). Based on the EC (ds/m) of all soil samples in the region, 36% of the samples were categorized as slightly saline, 22% as moderately saline, 33% as strongly saline, and 7% as very strongly saline (**Table 2**).

Table 2. Soil salinity categorization based on Electrical Conductivity (EC)

Soil EC	Salinity level	No. of samples
EC < 0.5	Normal	0
0.51 – 2	Slightly saline	40 / 36.6%
2.1 – 4	Moderately saline	24 / 22%
4.1 – 10	Strongly saline	37 / 34%
EC > 10	Very strongly saline	8 / 7.3%

3.1.2 Remote sensing data

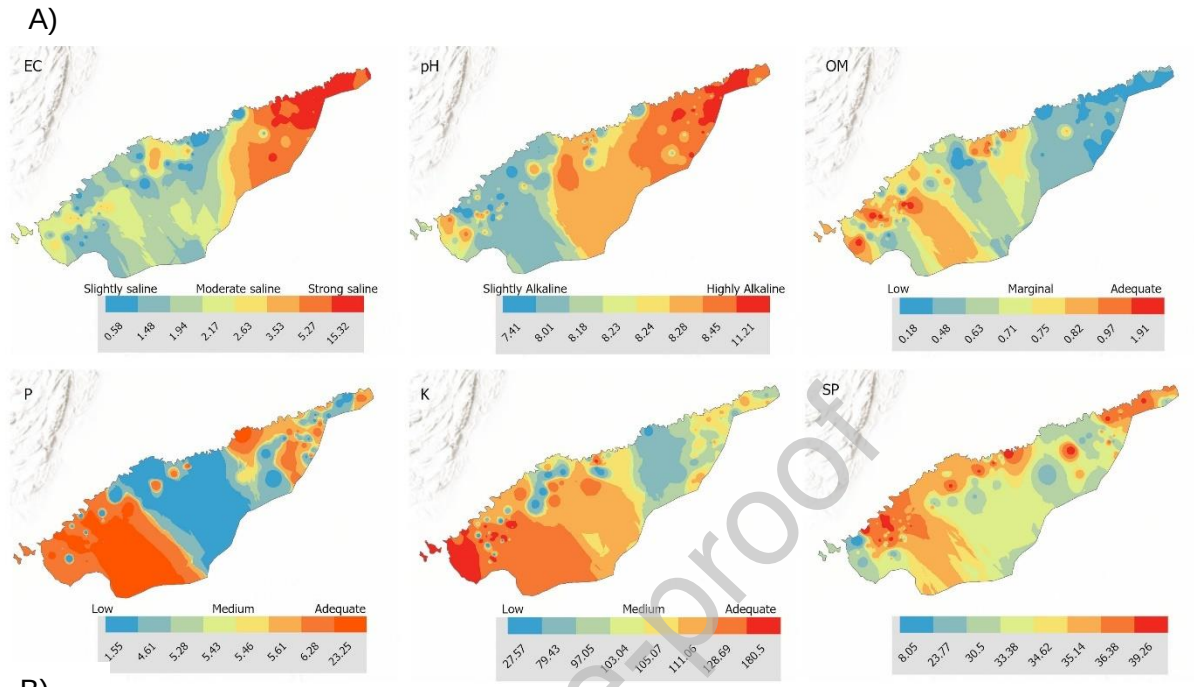
A range of remote sensing data have been utilized in previous research for soil salinity monitoring and prediction. Spectral information derived from Landsat with 30m spatial resolution and Sentinel with 10–20m spatial resolution has been proven to have a good correlation with EC measurements [53]. However, physicochemical differences in soil

properties necessitate reevaluation in different regions. The surface reflectance product of the Copernicus satellite, Sentinel 2 MSI, was retrieved from the Google Earth Engine (GEE) in a similar time frame of soil sample collection to map and extract index values for all indices. The median image of the entire year, as suggested by Lobell, Lesch [54] was selected to provide more stability in the average reflectance of the soil. Three vegetation indices, namely, the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Chlorophyll Vegetation Index (CVI), were used as vegetation predictors. Moreover, seven soil salinity indices derived from the red, green, blue, near infrared (NIR), and red edge bands of Sentinel 2 were calculated **(Table 3, Fig 2B**). The salinity index I (SI-1) was derived from the blue (B2) and red (B4) wavelengths, and the salinity index II (SI-2) was derived from the green (B3) and red (B4) wavelength bands of the Sentinel 2 MSI. Similarly, salinity index III (SI-3), salinity index IV (SI-4), and salinity index V (SI-5) were derived from the green, red, and NIR (B8) wavelength bands of the Sentinel-2 MSI with a 10 m spatial resolution. Furthermore, salinity index-VI (SI-6) added red edge 4 (B8A) along with the green and red wavelength bands. The Canopy Response Salinity Index (CRSI) is a measure of canopy reflectance in saline croplands [24], currently computed as salinity index-VII (SI-7) from red, green, blue, and red edge 4 wavelength bands of Sentinel 2 **(Table 3, Fig 2B**).

Moreover, Sentinel-1with 10 m spatial resolution, characterized by C-band SAR, was also retrieved from the cloud computing platform of GEE to extract single co-polarization band VV and Dual band cross polarization VH with good penetration ability to the underground surface (**Table 3, Fig 2B**). Synthetic Aperture Radar (SAR) polarization properties have been widely used in various soil monitoring phenomenon, such as soil moisture [55], soil texture [56], and soil salinity analysis [57]. The backscatter properties of SAR are highly related to several soil properties such as roughness and moisture, potentially leading to a significant relationship with soil salinity and crop vegetation; therefore, they are considered as potential predictors. Additionally, basic topographic factors, such as elevation, slope, and aspect (**Table 3, Fig 2B**), are also significant in determining the spatial patterns of soil properties. These were obtained from NASA SRTM 30m (SRTMGL1-003) in GEE and are considered to be significant predictors of EC measurements [58]. Along with these basic variables, the Topographic Wetness Index (TWI) was also calculated following the procedural steps of Buchanan, Fleming [59], using the following formula:

$$TWI = \ln\left(\frac{\alpha}{\tan b}\right)$$

Where α represents the upslope contributing area is draining through a point, $\tan b$ is the local slope in radians.



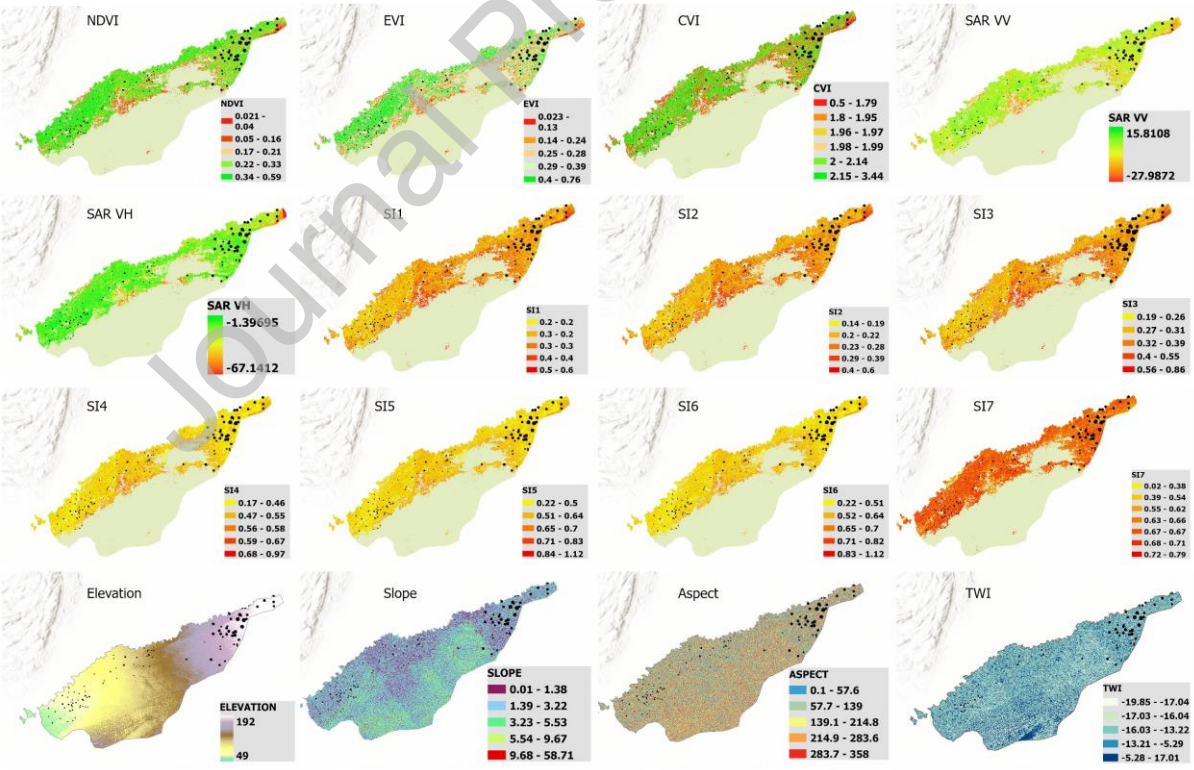


Fig 2 A) Spatial distribution of soil parameters (EC, pH, OM, P, K, and SP), **B)** spatial distribution of remote sensing indicators (vegetation indices (NDVI, EVI, CVI, SAR VV, SAR VH), salinity indices (SI1 – SI7), and topographical factors (elevation, slope, aspect, TWI).

Table 3. Categorization of all covariates with their abbreviations and formulas used for soil EC-based salinity predictions.

Category	Variables	Abb	Formula/units	Reference
Soil	Electrical Conductivity (dS/m) pH, Organic matter (%), Available Phosphorous (ppm), Extractable Potassium (ppm), Saturation (%)	EC pH, OM, P, K, Sp	----------	Soil survey and analysis
Vegetation Indices	Normalized Difference Vegetation Index	NDVI	$\frac{NIR - RED}{NIR + RED}$	[60]
	Enhanced Vegetation Index	EVI	$2.5 * \left(\left(\frac{REDEDGE - RED}{REDEDGE + 6 * RED - 7.5 * BLUE + 1}\right)\right)$	[61]
	Chlorophyll Vegetation Index	CVI	$\left(\frac{NIR}{GREEN}\right) * \left(\frac{RED}{GREEN}\right)$	[62]
Salinity Indices	Salinity Index-I	SI-1	$\sqrt{Blue * Red}$	[63]
	Salinity Index-II	SI-2	$\sqrt{Green * Red}$	[63]
	Salinity Index-III	SI-3	$\sqrt{Green^2 + Red^2}$	[64]
	Salinity Index-IV	SI-4	$\sqrt{Red^2 + NIR^2}$	[63]
	Salinity Index-V	SI-5	$\sqrt{Green^2 + Red^2 + NIR^2}$	[64]
	Salinity Index-VI	SI-6	$\sqrt{Green^2 + Red^2 + Rededge4^2}$	[65]
	Canopy Response Salinity Index (CRSI)	SI-7	$\sqrt{\frac{(Rededge4 \times Red) - (Green \times Blue)}{(Rededge4 \times Red) + (Green \times Blue)}}$	[66]
Synthetics Aperture Radar Backscatter Coefficients	vertical transmit/vertical receive	SAR VV		[67]
	vertical transmit/horizontal receive	SAR VH		[67]

Topographical / Terrain features	Elevation Slope Aspect Topographic Wetness Index	------- TWI	$\ln\left(\frac{\alpha}{\tan b}\right)$	[59]

3.2 Statistical analysis of all covariates

Initially, all covariates (soil and remote sensing indicators) were statistically analyzed to explore their descriptive characteristics. Subsequently, the data were divided into two groups based on the soil texture (loam and sandy loam). A non-parametric Tukey's test [68] was employed to examine the significant mean differences between the characteristics of loam and sandy loam soils. Pearson correlation and Principal Component Analysis [69] were employed to examine the significant ($p < 0.05$, $p < 0.01$, $p < 0.001$) linear relationship and dimensionality of all covariates and response variables.

3.3 Modeling soil salinity using machine and deep learning models

Classical machine learning models, such as Random Forest and Support Vector Machine, have already proven to be efficient in predicting soil salinity at various spatiotemporal scales [29, 70]. Deep learning is a specific approach of artificial intelligence that utilizes hierarchical structure of artificial neural networks. These are widely applied in the fields of image classification, language processing, speech recognition, and time series analysis [71]. Currently, deep learning architectures, including the feedforward neural network (FFNN), recurrent neural network (RNN), long short-term memory (LSTM), improved FFNN, improved LSTM, and ensemble (imp-FFNN + imp-LSTM), have been employed to predict EC-based soil salinity from 109 sampled locations in the middle-lower Indus basin of Pakistan. 21 multisource covariates were used as predictors of soil salinity in the region. However, selecting the most relevant predictors was an essential part of the analysis. Therefore, a wrapper-based feature-selection technique was employed to determine the most relevant predictors. Subsequently, the top selected features were input into six deep learning models to predict EC-based salinity measurements. The dataset was randomly split into 75% training (n = 81) and 25% testing (n = 28) datasets. Additionally, a 10% validation split was set during model training to estimate model performance through training and validation loss assessment using the mean squared error (MSE). The models were trained over a range of 100–200 epochs, with batch sizes ranging from 32 to 64. All DL models were trained in

Python 3 installed in a Core i7-12700H CPU with 24GB RAM using T4GPU computational capabilities. The libraries used included *Scikit learn*, *Keras, Pytorch,* and *SHAP*.

3.3.1 Covariates Selection (Recursive Feature Elimination with Random Forest)

Initially, to select the most appropriate covariates for the model input, the Recursive Feature Elimination (RFE) feature selection method was used along with Random Forest (RF) [72]. RF is a non-parametric ensemble machine learning model based on the concepts of bagging and bootstrap sampling. The ensemble nature of the model facilitates accurate feature engineering to extract the most relevant features from a dataset [73]. This is combined with recursive feature elimination, which works by recursive removal of the least important feature from the dataset until the desired subset is achieved. Currently, RFE is employed by fitting the Random Forest model with an input of 21 multisource covariates or features (**Table 3**). Feature importance was evaluated using a defined metric called Gini impurity. To obtain the maximum variables, a minimum threshold of 10 selected features (n =10) was set. The iterative process continued until the desired number of features was achieved. The resultant RFE-RF feature importance provides insight into the ranked scores of all features. RFE-RF is an advantageous wrapper-based method in several aspects, such as handling high-dimensional data and mitigating overfitting by reducing irrelevant and redundant information [74].

3.3.2 Feedforward Neural Network (FFNN)

The Feedforward Neural network (FFNN) is a type of ANN architecture that comprises interconnected nodes with unidirectional flow of information from the input layer through hidden to output layers [75]. In the hidden layers, output of each node (h_i) is computed using weighted sum of previous layer, ($\sum_{j=1}^{n} w_{ij}\, x_j + b_i$) followed by an activation function ($f.$) , which introduces non-linearity in the model (**eq. 1**). The common activation functions employed are sigmoid, tangent, and ReLU. Hence, nonlinear transformation allows the network to learn complex spatial and temporal relationships [76].

$$\boldsymbol{h_i = f.\left(\sum_{j=1}^{n} w_{ij}\, x_j + b_i\right) \ldots\ldots\ldots 1)}$$

Currently, our simple FFNN model consists of three fully connected dense layers with two hidden layers of 32 neurons each, utilizing the ReLU activation function. The last dense layer with a single neuron corresponds to the FFNN model output. Adam optimizer was used to

compile the model with a MSE loss function. The model was trained over 100 epochs with a batch size of 32. Furthermore, the improved FFNN architecture was optimized with three dense layers, the 1st layer comprised 100 neurons, followed by a dropout layer with a 0.2 dropout rate for regularization. Subsequently, the 2nd layer with 50 neurons was also followed by a dropout layer to mitigate overfitting, and the model was concluded with a single-unit dense layer with linear activation (**Fig 3**). The model was compiled using Adam, optimized with a learning rate of 0.001 and MSE loss function. The improved model was trained over 200 epochs with a batch size of 64 and 10% validation split.

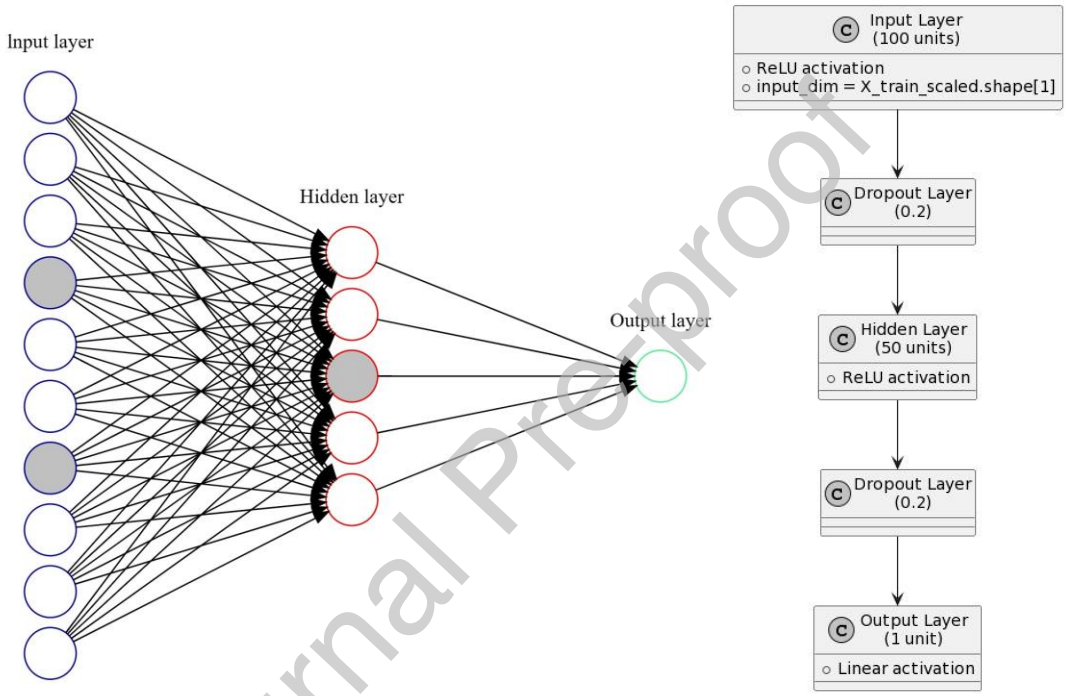


Fig 3 Improved FFNN architecture, input layer (100 neurons scaled on 10 with 0.2 drop out in grey shaded), hidden layer (50 neurons scaled on 5 with 0.2 drop out in grey shaded), and single unit output layer.

3.3.3 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN) are a type of advanced neural network architecture specifically designed for sequential data handling, adopting the knowledge of past steps. The feedback signal of the RNN from output to input enables the model to learn from past information [77]. Unlike FFNN, recurrent neural networks possess unique temporal behavior with directed graph connections between nodes that have a memory to retain previous

information. The key components of a simple RNN include the input sequence, recurrent connection, hidden state, weight metrics, bias term, activation function, and output [78]. Typically, RNNs are configured with uni or bidirectional layers, whereas uni or a simple RNN encounters the obstacle of vanishing gradients [79]. The gradients exponentially decrease as they propagate backward over time during training. This impedes the network's ability to acquire knowledge from lengthy sequences, as prior layers might receive inconsequential gradient updates, resulting in challenges in comprehending and preserving pertinent relationships [71]. Moreover, in a classical RNN, forward processing potentially overlooks the temporal features extracted from reverse time direction [80]. Currently, our simple RNN was defined using 'Sequential' API from Keras with single RNN layer of 50 units and a dense layer with single unit. The model was trained using Adam optimizer at a learning rate of 0.001. The model was optimized using the MSE loss for 100 epochs at a batch size of 32. Moreover, a 10% validation split was used to monitor the model training with training and validation loss metrics over epochs.

3.3.4 Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) is an advanced variant of simple RNN that overcomes the RNN limitations of vanishing gradients and improves memory ability over long sequences [79]. The sophisticated architecture of LSTM features the memory cells, an input gate (i_t), a forget gate (f_t), and output gates (O_t) (**Fig 4a**).This mechanism enables the LSTM to retain or discard information. Multiple gates and memory cells enable users to uphold information across sequences. The mechanism of LSTM multiple gates is as follows. Firstly, the input gate (i_t), which decides which information to be kept in a cell state and consists of sigmoid activation function (σ) (**eq. 2**), followed by the forget gate (f_t) deciding which information to be discarded from cell state (**eq. 3**). The procedure follows by computing the candidate values $C{\sim}t$ to be added in cell state C_t, which is computed from (f_t) and (i_t).

$$\text{(Input gate)}\ \boldsymbol{i_t = \sigma\,(W_i.[h_{t-1}, X_t] + b_i} \ldots\ldots\ldots\ldots \mathbf{2})$$

$$\text{(Forget gate)}\ \boldsymbol{f_t = \sigma\,(W_f.[h_{t-1}, X_t] + b_f} \ldots\ldots \mathbf{3})$$

$$\text{(Update cell state)}\ \boldsymbol{C{\sim}t = \tanh\,(W_c.[h_{t-1}, X_t] + b_c} \ldots\ldots \mathbf{4})$$

$$\boldsymbol{C_t = f_t.C_{t-1} + i_t.C{\sim}t} \ldots\ldots\ldots \mathbf{5})$$

Afterwards, the output gate O_t, picks the output through sigmoid σ, W_o and b_o are parameters to be trained (**eq. 6**) and new hidden state h_t is computed by **eq. 7**.

$$\text{(Output gate) } \boldsymbol{O_t = \sigma(W_o.[h_{t-1}, X_t] + b_o} \ldots\ldots \mathbf{6}$$

$$\text{(Hidden state) } \boldsymbol{h_t = O_t.\tanh(C_t)} \ldots\ldots \mathbf{7}$$

Overall, X_t is the current input, h_{t-1} is the previous hidden gate, and W and b terms define the weights and bias in respective gates.

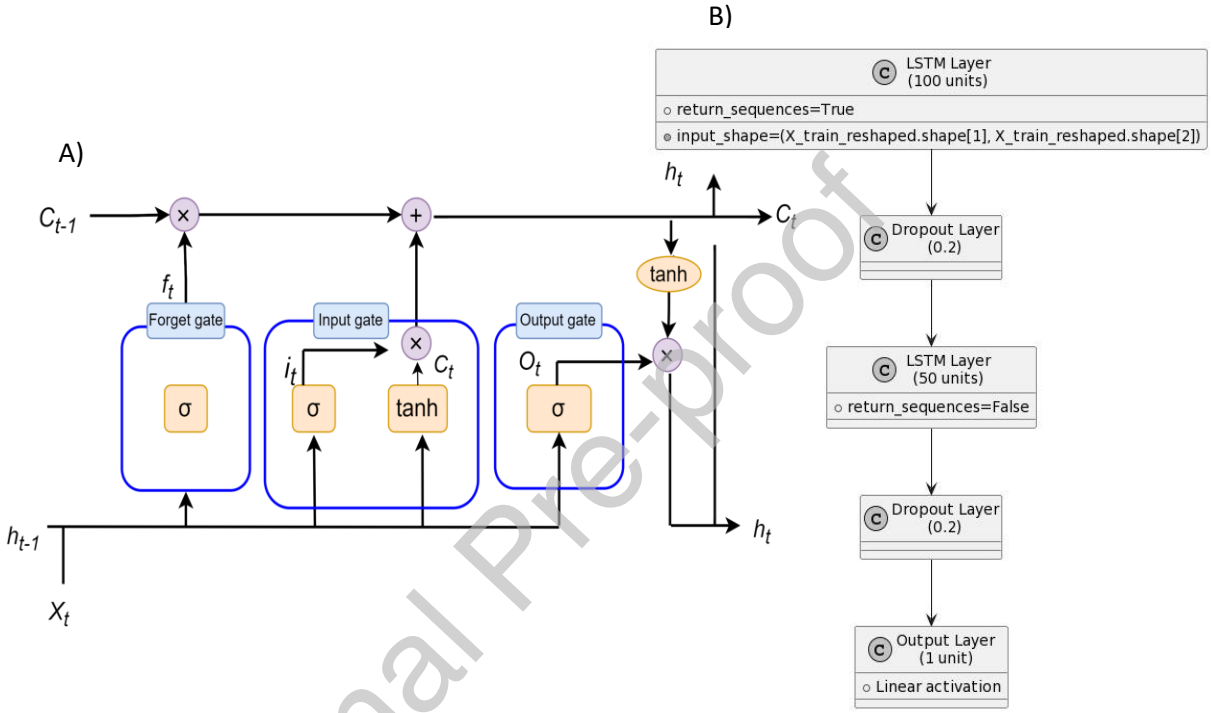


Fig 4 A) Generic LSTM architecture presenting the flow of information from input gate, forget gate, and output gate. **B)** Improved sequential LSTM architecture with 2 LSTM layers of 100 and 50 neurons with 0.2 dropout followed by linear output layer.

Currently, the simple architecture of LSTM was characterized by an input layer with shaped sequences followed by a 50-units LSTM layer to capture the spatiotemporal dependencies. Subsequently, a dense single-unit output layer compiled the model output. Like simple RNN, LSTM model was also compiled using the Adam optimizer with an MSE loss function over 100 epochs. To enhance the prediction accuracy, the LSTM model was further optimized and named 'imp-LSTM'. It comprised of two LSTM layers stacked with varying units. The 1st LSTM layer consisted of 100 units with a 'true' return sequence, while the subsequent layer had 50 units with a 'false' return sequence. Additionally, the model incorporated dropout regularization technique to mitigate overfitting after each LSTM layer. Finally, a single-unit dense layer with linear activation is employed to generate the predicted outputs (**Fig 4**).

3.3. SHapely Additive exPlanations (SHAP) model interpretation

The black-box nature of machine and deep learning models introduces uncertainty and complexity in interpretation. Multiple model interpretation algorithms are used in literature such as permutation importance [81], LIME [82], and SHAP. The SHapely Additive exPlanations (SHAP) was developed by Lundberg, Nair [83], and based on the game theory [84]. It provides more in-depth insight into the model with a quantitative contribution (positive or negative) of each parameter in making predictions [84, 85]. For every predicted sample of soil EC ($\widehat{\boldsymbol{y}}_{\boldsymbol{i}}$), the model generated a predictive value, with a SHAP value $\boldsymbol{f}(\boldsymbol{x}_{\boldsymbol{ij}})$ representing the contribution of each covariate.

$$\widehat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{Y}_{\boldsymbol{base}} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \boldsymbol{f}\left(\boldsymbol{x}_{\boldsymbol{ij}}\right) \ldots\ldots\ldots \mathbf{8})$$

Where, the mean of target variable for all samples is presented by Y_{base}, if, SHAP $f(x_{ij}) > 0$ for a feature, indicates that prediction value is improved by that feature. The higher the value of $f(x_{ij})$, the greater its impact on the model. Another advantage of the SHAP interpretation is its negative and positive value for each feature that determines the nature of the impact. Currently, SHAP is employed to address the uncertainty of the model and the associated features or covariates' impact on salinity prediction.

3.3.6 Model Evaluation metrics.

After the successful implementation of the proposed deep learning models, the performance was evaluated based on the coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Nash-Sutcliffe efficiency (NSE) (**Table 4**).

Table 4. Evaluation metrics used for models' interpretation

Indicator	Equation	Range	Best fit	Eq no
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(EC_{pred} - \overline{EC}_{obs}\right)^2}{\sum_{i=1}^{n}\left(EC_{obs} - \overline{EC}_{obs}\right)^2}$	0 to +1	+1	**8**
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(EC_{obs} - \widehat{EC}_{pred}\right)^2$	0 to ∞	0	**9**
$RMSE$	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(EC_{obs} - \widehat{EC}_{pred}\right)^2}$	0 to ∞	0	**10**
MAE	$MAE = \frac{\sum_{i=1}^{n}\left\lvert EC_{pred} - EC_{obs}\right\rvert}{n}$	0 to ∞	0	**11**

$MAPE$	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{EC_{obs} - EC_{pred}}{EC_{obs}}\right\rvert$	0 to ∞	0	**12**
NSE	$NSE = 1 - \frac{\sum_{i=1}^{n}\left(EC_{obs} - \widehat{EC}_{pred}\right)^2}{\sum_{i=1}^{n}\left(EC_{obs} - \overline{EC}_{obs}\right)^2}$	- ∞ to +1	+1	**13**

4. Results

4.1 Statistical analysis of remote sensing and soil variables

Descriptive statistics of all soil samples revealed that minimum EC for the loamy soil was 0.57dS/m and maximum was 11.5dS/m, with a mean EC of 3.5dS/m ± 2.82 SD. This was followed by a significantly ($p < 0.001$) high mean EC of 8.5 dS/m ± 9.39 for the sandy loam (**Fig 5**). Similarly, the minimum pH for the loamy soil samples was found to be 7.4 and maximum of 11.9 with a mean of 8.34 ± 0.46SD. Sandy loam soil showed significantly ($p < 0.001$) high alkalinity with a high mean pH of 9 ± 0.94SD. The minimum organic matter content for the loamy soil ranged from a minimum of 0.17% to a maximum of 1.91%, with a mean of 0.71 ± 0.33SD. For the sandy loam soil, the mean organic matter was found to be significantly ($p < 0.01$) less than 0.51% ± 0.23SD than the loamy soil (**Fig 5**). Overall, the average organic matter in both types of soil was found to be in a low threshold i.e., less than 0.86. For other soil nutrients, the mean of available phosphorous (P) in the sandy loam was found to be low i.e., 5.58 (ppm) ± 2.0SD, and low to medium 7.01(ppm) ± 6.85SD for loamy soil. The mean available potassium (K) for both soil types was found to be in the medium range, 129 (ppm) ± 65SD for loam and 110.8 (ppm) ± 50.5SD for sandy loam. The mean saturation percentage (SP) of loam soil was 35% ± 4.63 and sandy loam was 33% ± 4.8. The mean difference analysis for the vegetation indices (NDVI, EVI, and CVI) revealed lower mean values of 0.21, 0.25, and 1.83 for sandy loam (**Appendix Table 1, Fig 5**). Furthermore, the salinity indices (SI-1 to SI-7) showed higher mean index values for sandy loam than for loamy soil. For sandy loam soil, the SI-1 mean value 0.23 ± 0.04SD was higher than loamy soil i.e., 0.20 ± 0.05SD. Similarly, SI-3 also exhibited a higher mean 0.36 ± 0.08SD compared with loam soil 0.32 ± 0.08SD. The mean index value of SI-7 (canopy response salinity index) was highest 0.66 ± 0.06SD for sandy loam than 0.64 ± 0.04SD for loam soil. Furthermore, for SAR VV, the mean backscatter coefficient for sandy loam was -11.73 ± 3.6SD, while for loamy soil was -10.81 ± 3.2SD. Similarly, for SAR VH, the mean of backscatter coefficient for sandy loam was -18.9 ± 3.7SD followed by -17.6 ± 3.6SD. For

topographical factors, a high mean difference was observed for elevation, with a mean sandy loam soil presence of 143 meters ± 28SD than 111meters ± 31SD for loamy soil (**Appendix Table 1, Fig 5**).

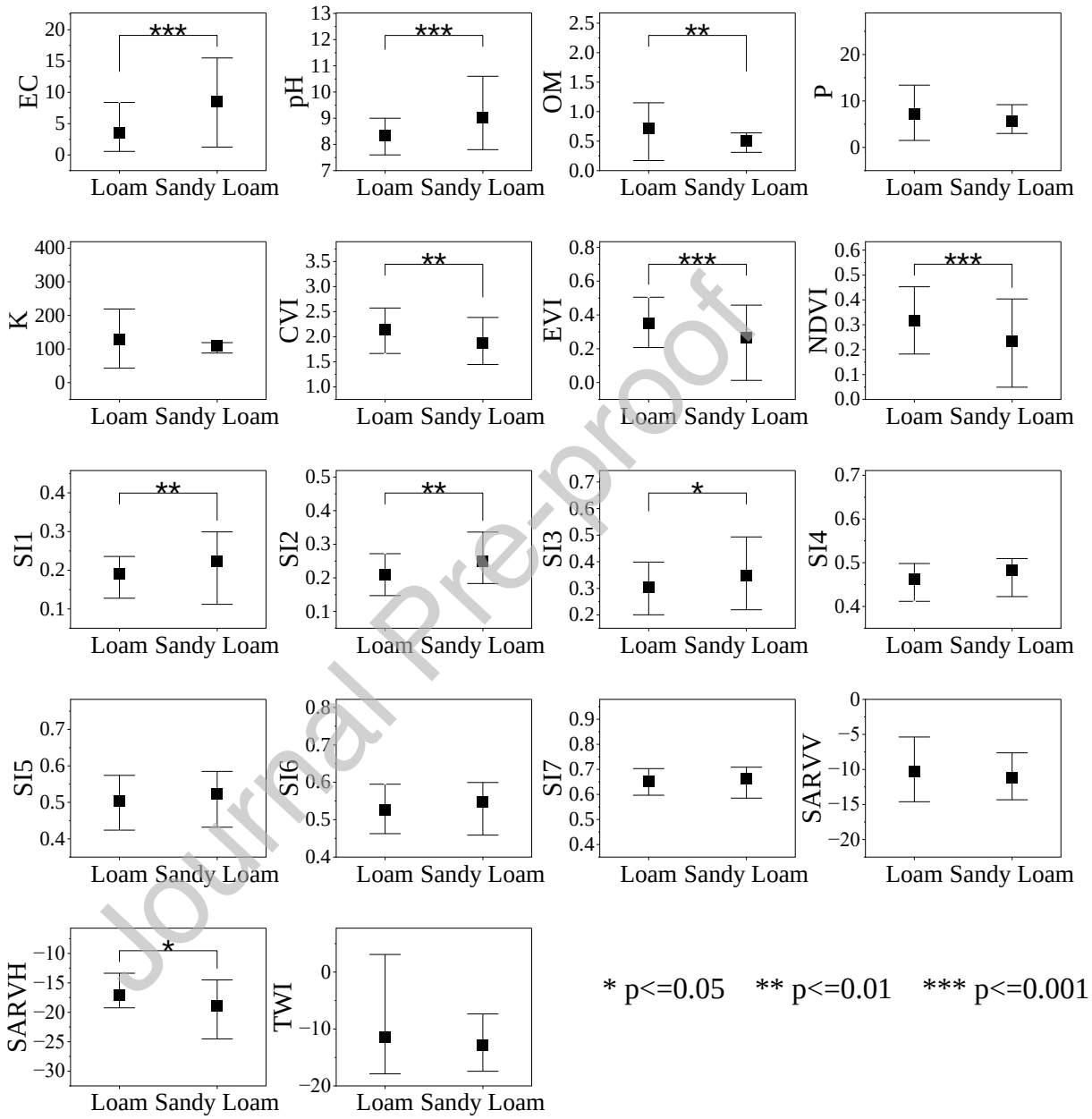


Fig 5 The output of the Tukey test for all covariates in loam and sandy loam soils.

Furthermore, Pearson correlation analysis between all covariates (soil and remote sensing) and the explanatory variable (EC) revealed a significant ($p < 0.05$, $p < 0.01$, $p < 0.001$) relationship between them. EC was found to exhibit a highly significant ($p < 0.001$) positive

correlation r = 0.7 with pH and r = 0.8 with elevation. In contrast, the EC revealed a significant ($p<0.01$) negative correlation r = -0.3 with organic matter and a significant ($p < 0.05$) low negative correlation r = -0.2 with available P and K. Among the remote sensing salinity indices, EC exhibited a significant ($p < 0.001$, $p < 0.01$) positive correlation, ranging from a minimum r = 0.3 with SI-VII to a maximum r = 0.6 with SI-I and SI-III. Among the topographical factors, EC exhibited a non-significant and weak relationship with slope (r = -0.01), aspect (r = 0.11), and TWI (r = -0.1). Among the vegetation indices, EC exhibited a significant negative correlation ($p < 0.001$) negative correlation with r = -0.6 with NDVI, EVI, and CVI. Backscatter polarization from SAR, such as VV and VH, also exhibited a negative correlation of r = -0.3 with EC. Moreover, all salinity indices (SI-1 to SI-6) and vegetation indices (NDVI, EVI, and CVI) were found to exhibit a positive correlation with each other. However, SI-7 (Canopy response salinity index) found to exhibit a weak correlation of 0.1 to 0.2 with other salinity indices.

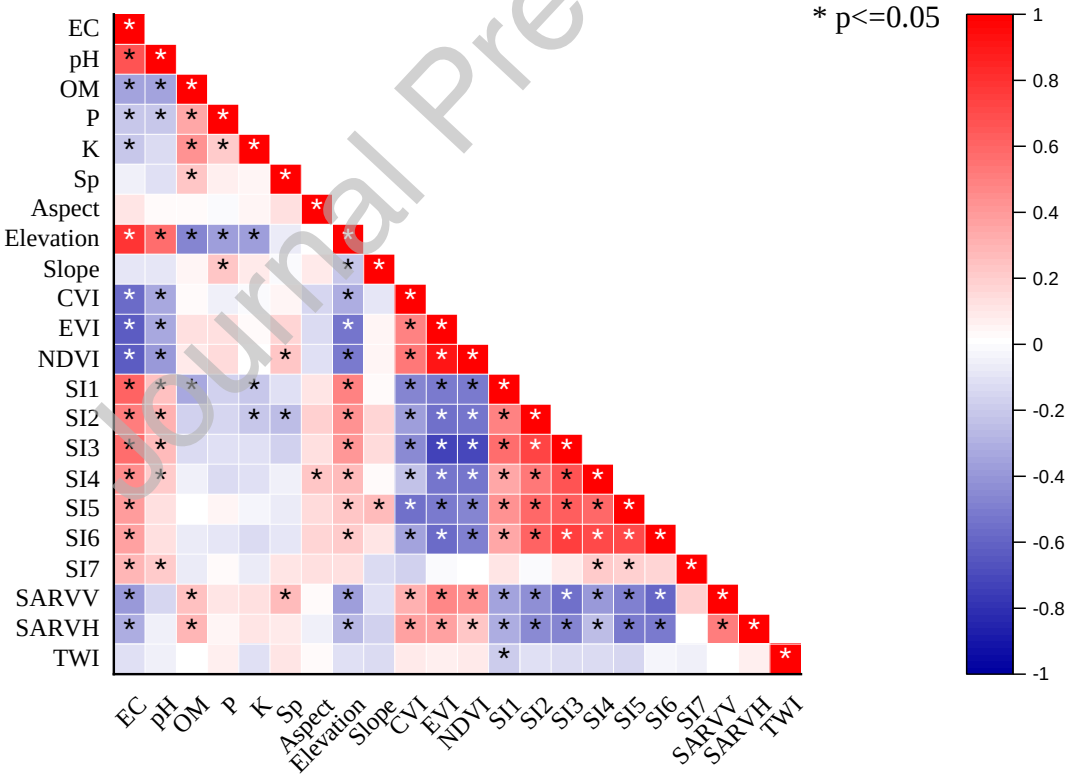


Fig 6 Pearson correlation coefficients among all variables (soil and remote sensing)

4.2 Principal Component Analysis

The PCA analysis of representative soil samples showed a high dimensionality in the data, with five PCs above one eigenvalue explaining a total variance of 63.7% in the data. The 3D PCA presentation showed that PC1 accounted for 33.6% of the explained variance, followed by 11% PC2, and 7.5% PC3, with a cumulative explained variance of 52% (**Fig 7, Appendix Table 2**). Furthermore, EC, pH, and elevation exhibited a strong positive correlation with each other, each showing a high positive loading in PC1, with coefficient values of 0.29, 0.17, and 0.24, respectively. PC1 was also positively influenced by SI-1 to SI-6, with a minimum coefficient of 0.25 for SI-1, 0.31 for SI-3. SI-7 (CRSI) significantly contributed to PC3, with a high positive loading and a coefficient of 0.46. Contrary, optical vegetation indices such as NDVI, EVI, and CVI, and radar backscatter coefficients such as SAR VV and SAR VH exhibited negative loadings in PC1. Other soil indicators, such as OM, P, and K, which were negatively correlated with EC, exhibited a high positive loading in PC2 and PC3 (**Fig 7, Appendix Table 3**). Furthermore, SP and aspect accounted for the high positive loading in PC3 (**Fig 7**, **Appendix Table 3**).

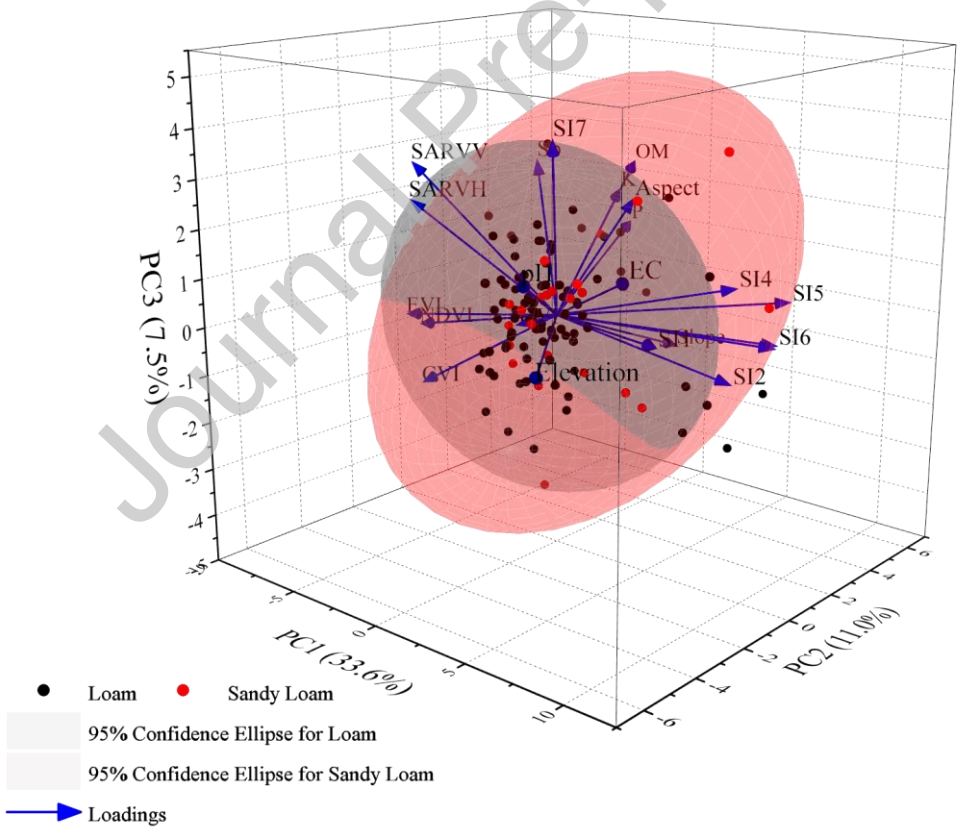


Fig 7 3-Dimensional presentation of PCA with variance % explained by each PC

4.3 Machine learning performance

4.3.1 Covariates selection through Recursive feature elimination

Initially, the RFE-RF selected the top ten important features (covariates) with no collinearity between them to be used as inputs in deep learning models for soil salinity prediction (**Fig 8**). The variable importance score by RF revealed the relative contribution of each feature to the EC-based measurements of soil salinity. A threshold of n = 10 features revealed that OM was the most effective predictor of soil salinity with a ranked value of 0.24, followed by SI-7 (CRSI) with a ranked value of 0.18. Subsequently, SP (0.18), pH (0.14), and elevation (0.11) from topographical factors also proved to be substantial predictors of the EC-based salinity measurements. Subsequently, SI-1, SI-5, SI-6, SAR VV, and NDVI were identified as significant inputs in the deep learning models for soil salinity prediction.

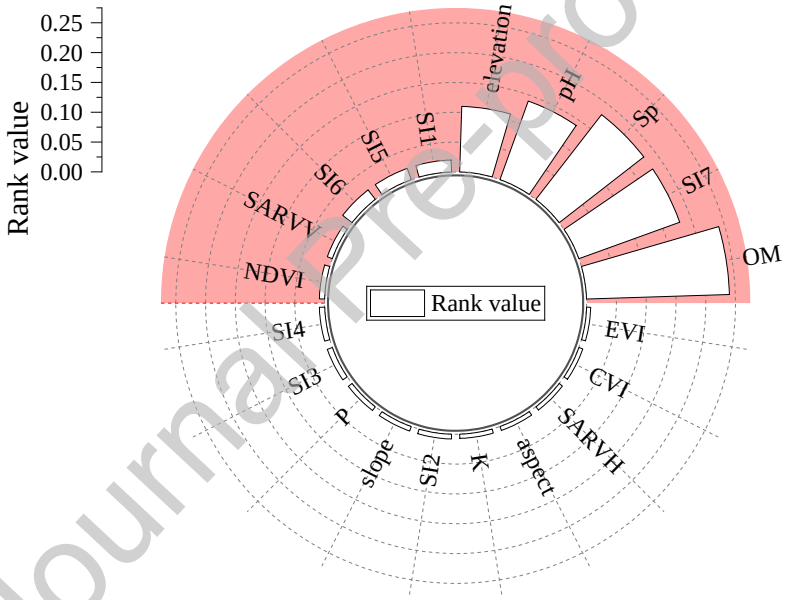


Fig 8 Feature (covariates) selection through variable importance and rank value from RFE-RF method (top 10 selected features with high ranks shaded in red)

4.3.2 Performance assessment of DL models in training and testing sets

The selected covariates from the RFE-RF were trained separately in the FFNN, RNN, LSTM, imp-FFNN, imp-LSTM, and ensemble (imp-FFNN + imp-LSTM). The evaluation metrics revealed that a simple FFNN with three fully connected layers without regularization achieved the highest R^2 = 0.889, lowest MSE = 1.66, RMSE = 1.288, MAE = 0.89, MAPE =

30.4%, and NSE = 0.888 in model training. However, it could not perform effectively on the testing set and was not considered the best performer. The ensemble of imp-FFNN and imp-LSTM performed competitively with a high R^2 = 0.886, MSE = 1.7, RMSE = 1.30, MAE = 1.02, MAPE = 36.4%, and NSE = 0.886 in training set. Afterwards, based on the performance score, imp-LSTM and imp-FFNN with denser and dropout layers with regularization showed a high R^2 = 0.883 and 0.870, MSE = 1.74 and 1.94, RMSE = 1.322 and 1.394, MAE = 1.032 and 1.109, MAPE= 35% and 40%, and NSE = 0.883 and 0.870 during the training phase (**Fig 10, Appendix Table 4**). A simple RNN with a dense layer of 50 neurons performed less with R^2 = 0.742, MSE = 3.84, RMSE = 1.961, MAE = 1.613, MAPE = 52.4%, and NSE = 0.742. Similarly, the simple LSTM architecture also showed the lowest performance, with a minimum R^2 = 0.709, MSE = 4.34, RMSE = 2.083, MAE = 1.698, MAPE = 52.44%, and NSE = 0.709 (**Fig 10, Appendix Table 4**). Overall, the FFNN model was trained with the highest accuracy, followed by the ensemble of imp-FFNN and imp-LSTM in the training sets. The simple architectures of LSTM and RNN did not perform well during the training of the model. It was improved by optimizing the architecture with additional dense and dropout layers, which helped mitigate overfitting and achieve more accurate predictions. The scatterplots with standard error bars between the original and predicted soil EC also revealed the outperformance of the ensemble of imp-FFNN and imp-LSTM with the least overfitting (**Fig 9**).

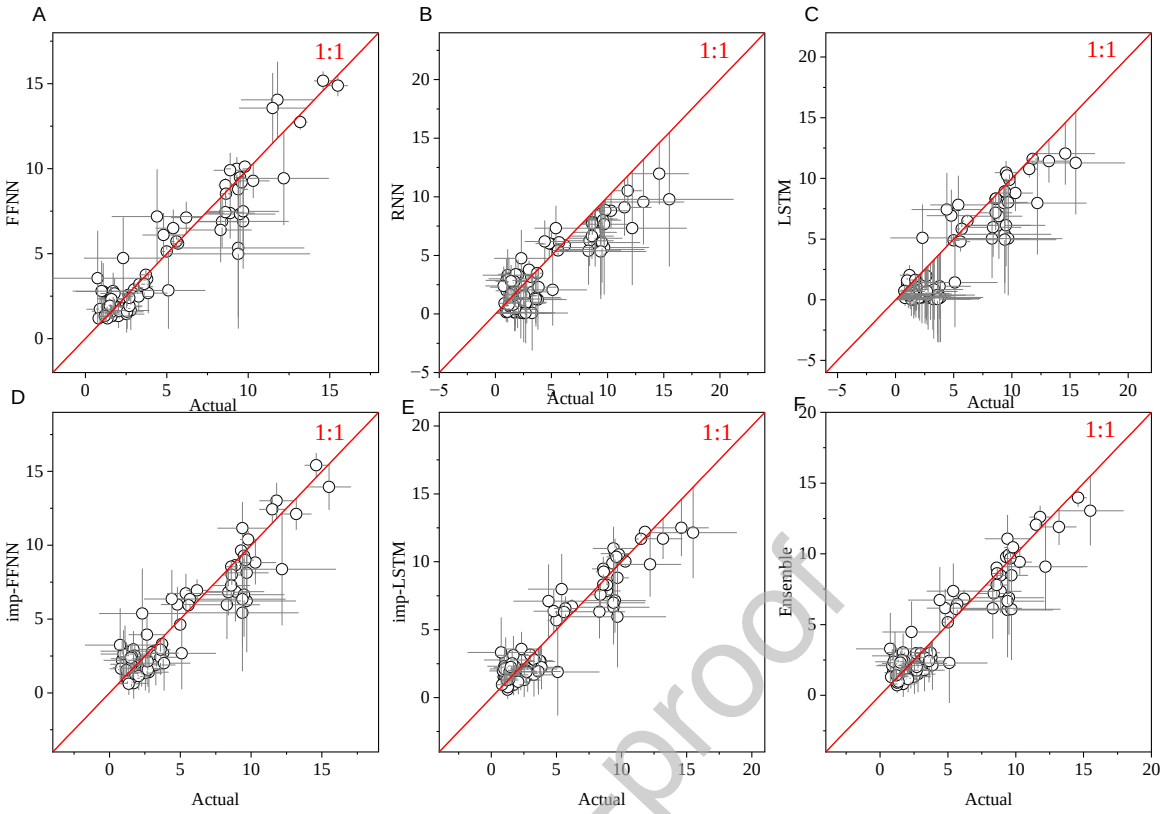


Fig 9 Scatterplots with St. Error between actual and predicted soil EC in all DL models **A.** FFNN, **B.** RNN, **C.** LSTM, **D.** Imp-FFNN, **E.** Imp-LST, **F.** Ensemble (train set predictions)

After successful model training, predictions on the test dataset revealed that the improved architectures of FFNN, LSTM, and their ensemble performed better for soil salinity predictions. Initially, a simple FFNN with three dense layers showed the lowest R^2 = 0.694, highest RMSE = 1.915, MAE = 1.330, MAPE = 43%, and NSE = 0.694. Simple RNN predictions on the test set showed better R^2 = 0.742, RMSE = 1.961, MAE = 1.613, MAPE = 51%, and NSE = 0.742. Subsequently, LSTM, an advanced variant of RNN, revealed R^2 = 0.713, RMSE = 1.855, MAE = 1.613, MAPE = 57.4%, and NSE = 0.713 for the EC-based salinity predictions (**Fig 10, Appendix Table 4**). Furthermore, the improved FFNN optimized with regularization improved the predictions on the test sets with high R^2 = 0.731 compared to simple FFNN R^2 = 0.694, low RMSE = 1.796, MAE = 1.231, MAPE = 41%, and NSE = 0.731. Similarly, imp-LSTM also showed an improved and high R^2 = 0.824 compared with the simple LSTM R^2 of 0.713. The other metrics also revealed an improved performance, with a low RMSE = 1.453, MAE = 1.036, MAPE = 34%, and NSE = 0.824. Finally, the ensemble of imp-FFNN and imp-LSTM provided the highest R^2 and NSE = 0.841, and the

lowest RMSE and MAE = 1.382 and 1.014, respectively (**Fig 10, Appendix Table 4)**. Hence, based on the prediction performance, the capability of deep learning models is ranked in the following order: ensemble > imp-LSTM > RNN > imp-FFNN > LSTM > FFNN. Scatterplots of observed predicted EC from all deep learning models also revealed the efficient performance of ensemble of imp-FFNN and LSTM achieving high accuracy (**Fig 11**). The predicted EC-based salinity maps generated by the improved DL models exhibited higher EC values in the northeastern region. The minimum EC predicted by the imp-FFNN was 0.727 dS/m, indicating slightly saline irrigated agricultural soil towards the central and southwestern parts of the region. In contrast, the maximum EC predicted by the imp-FFNN was 20.572 dS/m, indicating strongly saline soil. Similarly, imp-LSTM predicted a minimum EC of 0.604 dS/m and a maximum EC of 12.646 dS/m. The ensemble predicted a maximum EC of 16.86 dS/m which was closer to the original EC value of 15.31 dS/m in northeastern parts of the region (**Fig 12**).

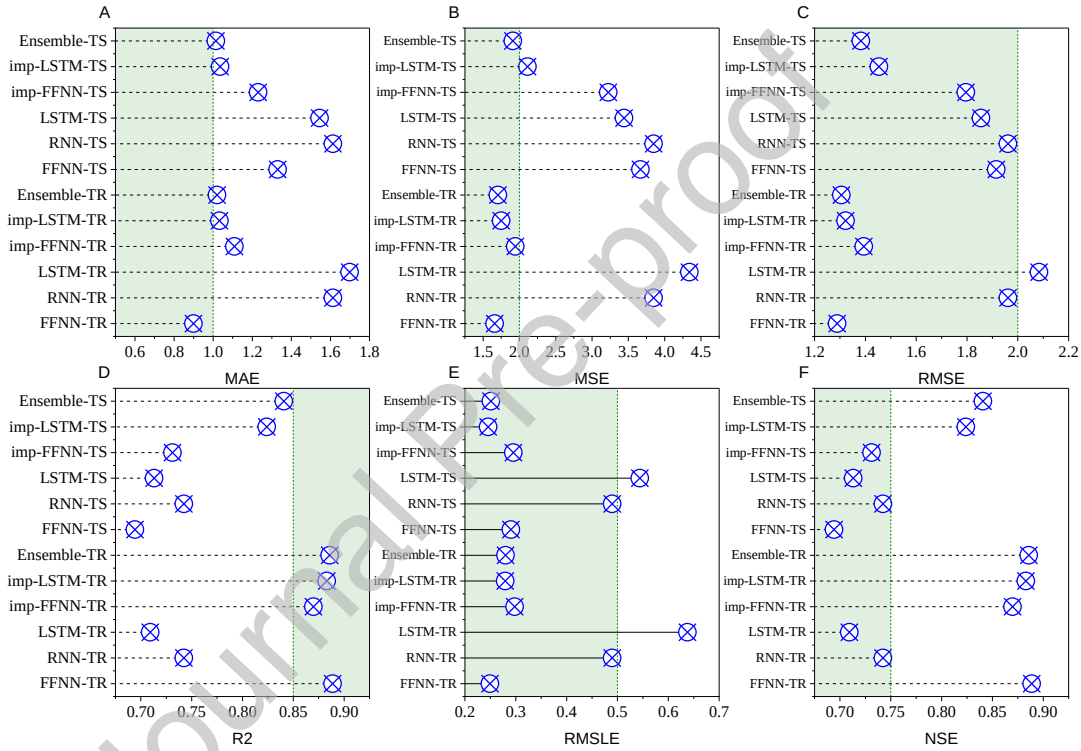


Fig 10. Evaluation metrics of deep learning models for soil EC prediction for training (TR) and testing (TS). A) Mean Absolute Error, B) Mean Squared Error, C) Root Mean Squared Error, D) R^2, E) Root Mean Squared Log Error, F) Nash-Sutcliffe Efficiency

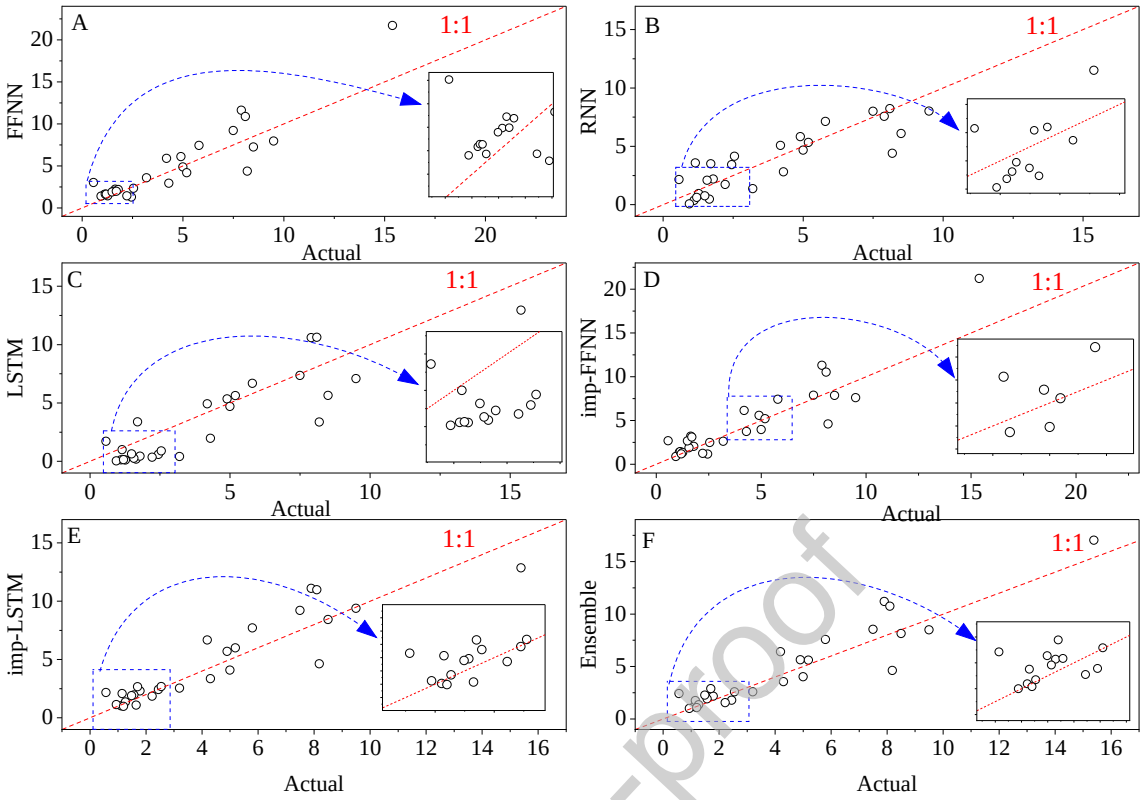


Fig 11 Scatterplots of the actual and predicted soil EC in all DL models **A.** FFNN, **B.** RNN, **C.** LSTM, **D.** Imp-FFNN, **E.** Imp-LST, **F.** Ensemble (test set predictions)

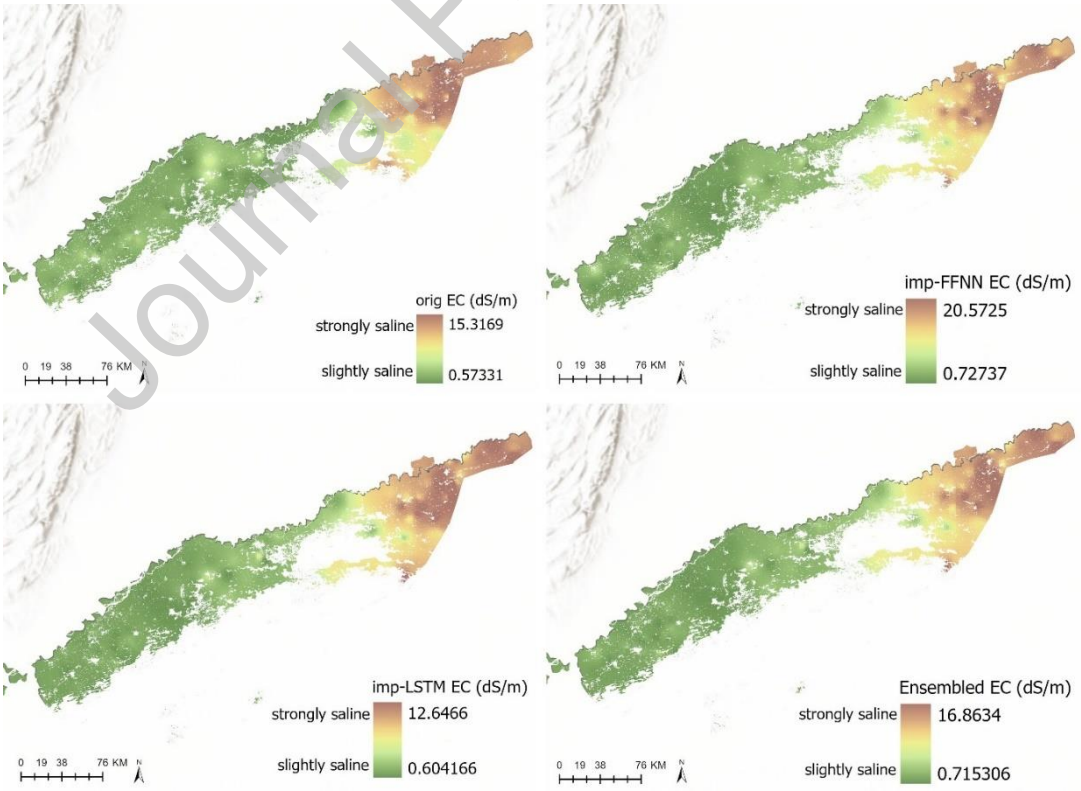


Fig 12 Soil salinity distribution based on predicted EC (dS/m) values from highly performed deep learning models

Furthermore, the generalization ability of the model was assessed based on the training and validation losses across epochs for each DL model (**Appendix Fig 1**). The consistent decrease in both the training and validation losses indicates improved model accuracy. The RNN and LSTM plots showed a gradual decline in training and validation losses over 100 epochs. The mean training losses for the RNN and LSTM were recorded as 15.31 and 16.57, respectively. The mean validation losses were also the highest for LSTM and RNN (**Appendix Table 5**). However, the FFNN showed improved model performance with the lowest mean training loss (8.89) and mean validation loss (11.29), indicating better generalization of the validation. Furthermore, the improved architecture of the FFNN and LSTM with regularization and early stopping showed a good convergence of training and validation loss from 30 to 80 epochs. The mean validation loss for imp-FFNN was 12.44, and for imp-LSTM was 14.73, suggesting better generalization and less or no overfitting (**Appendix Table 5**).

4.3.3 SHAP analysis for model's explanation

The feature summary plots from the SHAP analysis for imp-FFNN and imp-LSTM for training and test datasets reflect a straightforward relationship between multi-source covariates and predicted measurements of soil salinity. If the SHAP value is greater than 0, the covariate has a positive impact on the predictions. If the SHAP value is less than zero, the covariates have a negative impact on the prediction. Moreover, the colored bar from blue to red indicates a low-to-high intensity of impact. Currently, elevation is found to be the strongest determinant of soil EC in both models, with the highest contribution to model predictions. It exhibits spatially varied positive and negative impacts. Overall, the elevation of the study area was high in the northeast, with high EC values, and low in the southwest, with lower EC values. However, in few places in the central plain area of the region, high salinity was observed and predicted. Moreover, NDVI was found to be the 2nd and 3rd important contributor in the imp-FFNN and imp-LSTM models **(Fig 12).** The SHAP values for NDVI ranged between -2 to maximum +4, indicating both negative and positive impacts on predictions. High NDVI with positive SHAP values increased the salinity prediction, whereas NDVI with negative SHAP values decreased the predictions.

Similarly, SARVV was found to have a negative impact on the predictions of both models. On average, it was found to be the 6th and 7th significant contributor in the imp-LSTM and imp-FFNN models, respectively. Among the soil parameters, pH had a positive influence on soil salinity predictions with 2nd important contributor in imp-LSTM and 3rd in imp-FFNN. Among the salinity indices, SI1 and SI7 (canopy response salinity index) had a more positive

impact and 4th and 5th important contributors in both improved deep learning architectures. SI5 and SI6 had the least impact on soil EC predictions and ranked last in both models. Moreover, OM and SP were found to have low and varied impacts on soil EC predictions **(Fig 13)**.

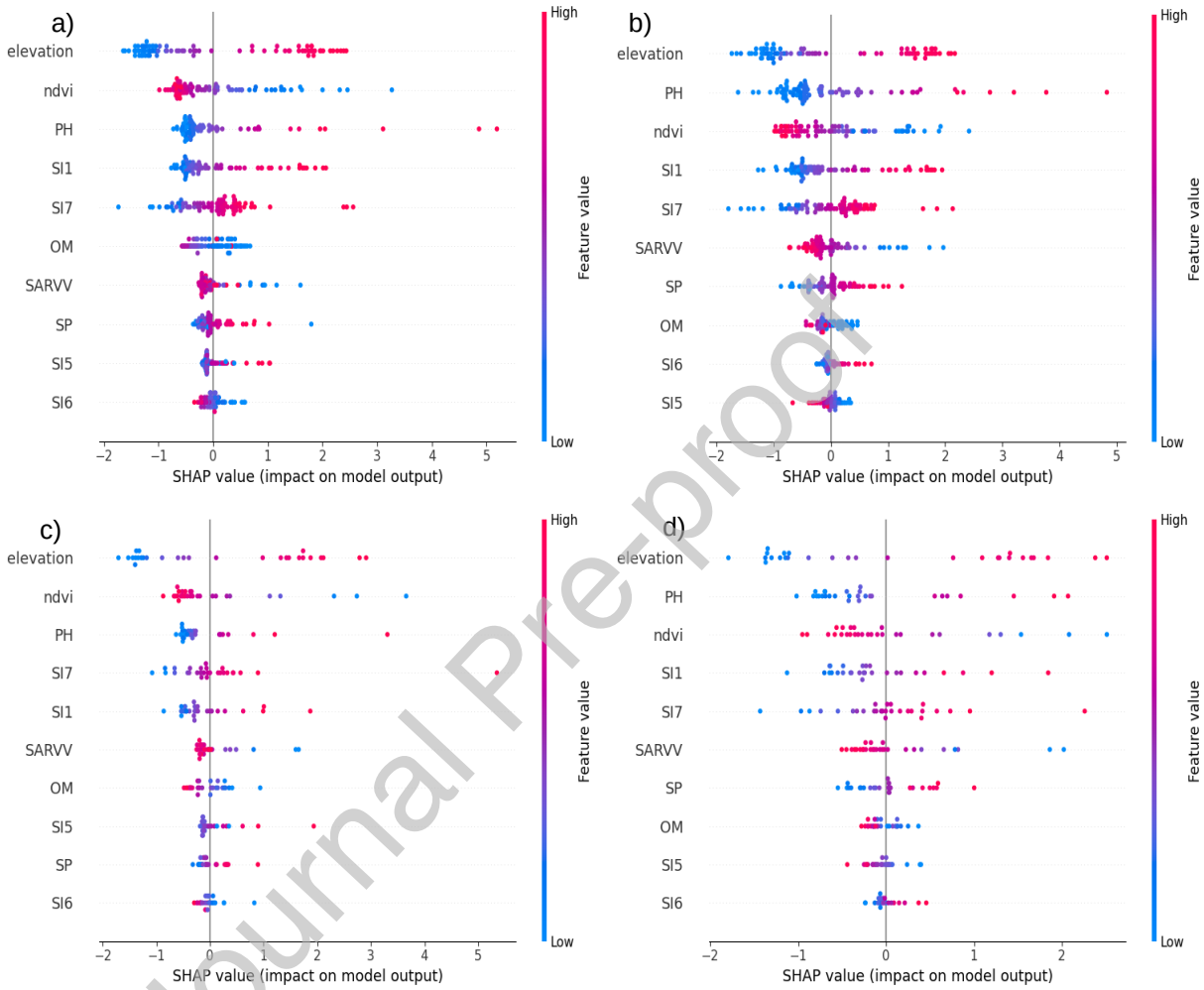


Fig 13: **a)** SHAP model impact for imp-FFNN (train), **b)** SHAP model impact for imp-LSTM (train), **c)** SHAP model impact for imp-FFNN (test), and **d)** SHAP model impact for imp-LSTM (test).

4. Discussion

4.1 Drivers and impacts of soil salinity in arid irrigated regions

Soil salinity is a global land degradation problem, and none of the continental land is free from negative impacts [6, 86]. Insufficient rainfall, along with agricultural practices in arid and semi-arid regions, leads to the accumulation of salts on the soil surface [8, 87]. Indo-Gangetic plains of South Asia are highly affected by salinization, with 10×10^6 hectares

being highly affected by salinity problems [88]. Several primary and secondary drivers of soil salinity have been reported in the region due to climatic and geological variations, as well as land use practices [19, 89, 90]. For instance, water table rise, flooding and overirrigation, seeping, and silting are prominent reported causes of soil salinity in arid irrigated lands of Pakistan, with subsequent impacts on agricultural productivity [91]. Currently, soil samples analysis from lower Indus basin of the country revealed 59% low to moderate salinity (EC 0.5 – 4 dS/m) and 41.3% with strong salinity (EC > 4.1 dS/m) **(Table 2).** Overall, most of the soil salinity in the Indus Basin is inherited by the parent material and soil formation processes [41]. Additionally, overexploitation of poor-quality groundwater has also become a substantial cause of water and soil salinization in agricultural regions [92]. Previously, Latif and Ahmad [93] reported that unequal distribution and scarcity of canal water forced groundwater extraction, leading to salinity and sodicity problems in the irrigated Indus Basin of Pakistan. Nevertheless, Corwin and Scudiero [19] identified soil salinization in their research area as a consequence of rapid evapotranspiration. This process causes capillary rise in the soil, leading to the accumulation of salts at the surface of the soil. Hence, in arid and semi-arid agricultural regions, a high rate of evapotranspiration, coupled with variability in the irrigation fraction and leaching factor, contributes more to increasing the salt concentration within the root zone [91, 94]. Interestingly, fossil salts with ionic species, such as Na^{+}, Ca^{2+}, Cl^{-}, CO_3^{2-}, and HCO_3^{-}, deposited in the deep strata of the Indus Basin are pumped out by tube wells during irrigation and deposited as subsurface layers [40]. Hence, extensive expansion of irrigation practices in the region, drainage blockage, canal seepage, and field percolation have caused the groundwater table to rise near the soil surface [95]. This leads to waterlogging and soil salinity, especially in saline groundwater areas, as observed in the northeastern parts of the region (**Fig 2A**).

4.2 Relationship of soil parameters and remote sensing indices with soil salinity

The parent material of the study area is mixed calcareous alluvium characterized by calcium carbonate ($CaCO_3$) and calcium sulfate ($CaSO_4$), with high alkalinity (pH > 7.0) and less organic matter (**Fig 2A**) [96]. This is consistent with our research findings revealing alkalinity of the soil, with pH range from 7.4 to 11.9 and less mean organic matter (0.5% in sandy loam to 0.7% in loam) (**Fig 5**). Moreover, the negative correlation of organic matter, phosphorus, and potassium with soil EC (**Fig 6**) reflects the dynamic interaction between them. Alkaline soil (pH > 7.0) inhibits the decomposition of raw organic matter by microorganisms, resulting in lower accumulation in the soil. Additionally, the dominance of calcium ions restricts the

availability of other cations [27]. Moreover, the positive correlation between soil EC and pH (**Fig 6**) clearly suggests a dynamic interaction between pH and soil EC, which is more likely related to the abundant Ca^{+2} ions in the parent material. Hence, the pH, OM, and SP were selected using the RFE-RF method (**Fig 8**). Furthermore, the high coefficients of EC and pH in PC1 showed that they were closely related to each other. Interestingly, elevation from topographical factors was also found to exert a strong influence on soil EC levels (**Fig 7**). Topographic factors such as elevation, slope, topographic position, and wetness index affect water flow, soil deposition, and sedimentation, which have been considered significant predictors of soil salinity in previous research [58, 97, 98]. Among all, elevation is currently observed to have high impact on soil salinity (**Fig 13**). This aligns with the findings of Wang, Hu [34] and Yang, An [99], which explain the increase in soil salinity levels at higher elevations attributed to geological structure, poor drainage, and limited leaching. Along with topographical factors, remote-sensing-based vegetation and salinity indicators have also been found to be influential in mapping and predicting EC-based salinity measurements [100, 101]. For example, Ijaz, Ahmad [44] employed three remote sensing vegetation indices (NDVI, SAVI, and MSI) to map soil salinity with high accuracy in Pakistan. This is consistent with the findings of Pérez González, García Rodríguez [102], in which NDVI is found to be significantly related with soil EC [97]. Three closely related vegetation indices (NDVI, EVI, and CVI), with high negative coefficients in PC1 (**Fig 7**), exhibited a significant ($p<0.05$) negative correlation with soil EC (**Fig 6**). However, to minimize the collinearity effect, NDVI was chosen as a more significant predictor of soil salinity by recursive feature elimination (**Fig 8**). Furthermore, SHAP analysis provides a better understanding of the uncertainty of all the selected features used for model training [34]. Therefore, the SHAP model interpretation also proved that NDVI had a strong influence on soil EC prediction (**Fig 13**), consistent with previous findings [103, 104]. In addition to optical sensors, the inclusion of SAR VV and VH polarization provides more powerful opportunities for subsurface penetration for soil salinity analysis [105]. In accordance with the sensitivity of VV polarization for structural characteristics of vegetation and soil moisture detection [106], it was selected by RFE as an important predictor of soil salinity. Moreover, it was ranked 6^{th} in the SHAP interpretation, underscoring its significance in accurate EC prediction (**Fig 13**). This is also consistent with previous findings by Ma, Ding [67]. Furthermore, optical wavelength bands such as blue, green, red, NIR, and red edge have also been proven to be effective for soil salinity mapping and prediction [36, 101, 107]. Of the seven salinity indices, SI-1, SI-5, SI-6, and SI-7, with a significant ($p>0.05$) correlation with EC (**Fig 6**), were selected by RFE for training the deep

learning models (**Fig 8**). Particularly, aligned with the findings of Wang, Shi [38], CI7 (CRSI) proved to highly influence the EC predictions, and ranked among top five in deep learning models [65]. EC mapping based on model predictions offers valuable insight for identifying highly saline areas within the region (**Fig 12**). It can assist the stakeholders prioritize immediate agricultural land management, such as crop selection, optimizing irrigation practices, and soil reclamation efforts [108]. These measures can improve soil health in salinity affected areas, supporting agricultural sustainability and ensuring high grain yield in alignment with SDG-2.

4.3 Applicability and comparison of machine and deep learning method for soil salinity prediction

Several machine learning techniques ranging from linear regularization to hierarchical tree-based and ensemble models have been found to be useful in predicting EC-based soil salinity measurements [29, 43]. For instance, Random Forest is already proved to be superior in predicting soil salinity at varied spatiotemporal scale [38, 98, 107, 109]. Currently, wrapper-based RFE combined with RF provides the best subset of selected features that are less correlated with each other and lead to improved model performance. Hence, eliminating collinearity, prevented the models from overfitting leading to better generalization on unseen data [110]. Among the various applied deep learning architectures, previous research on soil salinity measurements has proven the high accuracy of ANN-MLP [111, 112]. Our study used three well-recognized deep learning architectures (FFNN, RNN, and LSTM) to provide accurate predictions of EC-based salinity measurements in the study area. The scalability, feature learning architecture, and improved handling of complex relationships for high-dimensional inputs make them a better choice for EC prediction [80, 113, 114]. For instance, Mohammadifar, Gholami [115] proved the accuracy of the 1DCNN-LSTM hybrid model architecture for predicting soil salinity for more than 400 sampled locations. Moreover, Boudibi, Sakaa [111] demonstrated an MLP neural network for accurate salinity prediction. Currently, the simple architecture of FFNN employing two dense layers with 32 neurons with ReLU activation enables it to learn complex relationships of multisource variables with EC measurements and achieves the highest training R^2 of 0.889. Previously, Zaki and Abdul-Aziz [116] also outlined the significance of the simple architecture of FFNN for computational efficiency in prediction research. However, the simplest architecture of the FFNN failed to achieve a high prediction accuracy on the test dataset, with a minimum R^2 of 0.694. For a robust comparison of the best prediction model, the simple RNN and LSTM achieved a better prediction in the test dataset with R^2 = 0.742 and 0.713, respectively (**Fig 10**). Furthermore,

the improved architecture of FFNN and LSTM by adding more dense layers and neurons [117], dropout regularization, and learning rate adjustments [118, 119] facilitated improved model training with the lowest training and validation loss and improved predictions on test datasets (R^2 = 0.731 and 0.824, respectively) (**Fig 10**). The use of 'Adam' optimizer facilitated efficient convergence by adjusting learning rates during model training over 200 epochs leading to a stable optimization [120]. Hence, deep learning models have been proven to efficiently incorporate the spatiotemporal dependencies of EC measurements with effective and accurate salinity predictions [121]. Comparative assessment of deep learning architectures revealed a notable performance by improved LSTM over improved FFNN and RNN in understanding intricate dependencies in data. Memory cell combinations and gating mechanisms enable LSTM to retain previous information, exhibiting high performance with accurate predictions [122]. However, the ensemble of FFNN and LSTM provided highly accurate predictions of soil salinity, achieving the highest R^2 = 0.841, and proved to be advantageous for real-time monitoring of soil phenomena at a good spatial scale. Soil salinization is a region-specific problem that varies according to the local physicochemical soil properties [123, 124]. Therefore, the performance of machine-learning models must be tested across different regions. For example, RF has proven to outperform other models in several soil salinity prediction studies [31, 35, 38, 125], whereas boosting regression outperformed RF in other studies [29, 37]. Hence, validating the recommended machine or deep learning model across diverse agroecological regions is crucial to enhance its generalizability and ensure reliable salinity predictions.

5. Modeling uncertainties and other limitations

Overall, our study provides a detailed assessment of EC-based soil salinity predictions, proving the efficiency of advanced deep learning architectures. Despite accurate predictions achieved by the ensemble of imp-FFNN and imp-LSTM, the data scarcity due to the small sample size might also affect the prediction accuracy, and the models struggled to capture the full spectrum of soil variability across the study area. However, this was mitigated by a stratified soil sampling approach, maximizing the representation of agricultural land use. Moreover, the sequential architecture of deep learning models was well designed with a cross-validation technique to capture temporal dependencies and address the vanishing gradient problem in the case of RNN and LSTM [111, 121]. However, uncertainties might also exist in capturing spatial dependencies owing to a chance of spatial autocorrelation within the input data. Moreover, the RF-based RFE provides a more robust selection of appropriate features to be used as input in deep learning models to overcome the noise in

model training. Nevertheless, advanced hybrid deep learning models can be used in the future to address noise and outliers. The SHAP kernel explainer also provided good interpretability of the improved architectures, providing a better explanation of the covariate's impact on salinity predictions. However, more sensitivity analysis tests can be employed to test the uncertainties in the model's explanations [115]. Field data collection is a laborious and time-consuming task; therefore, it can be supported by proven data augmentation methods for better optimization [34]. Shortwave and Thermal infrared wavelengths have also proven to be more sensitive to soil salinity predictions and can be incorporated in future research [126, 127]. Hence, improvements in the sample size, remote sensing resolution, and comparative assessment of more deep learning architectures with sensitivity analysis can be incorporated into future research on salinity prediction.

6. Conclusion

Soil salinization is a challenge in agricultural practice. The current study examined and predicted EC-based soil salinity by integrating multisource remote sensing and soil variables using advanced machine and deep learning models. Based on the major findings of this research, soil pH was found to be significantly positively correlated with EC, while OM was found to be significantly negatively correlated. Vegetation indices (NDVI, EVI, CVI, SARVV, and SARVH) were found to have significant negative and soil salinity indices (SI1 to SI7) have positive correlation with EC. Among the topographical factors, elevation was significantly positively correlated with EC. Facilitated by the RFE-RF feature selection method, the top ten selected features were input into deep learning models (FFNN, RNN, LSTM, imp-FFNN, imp-LSTM, and ensemble) to achieve accurate salinity predictions. The improved architectures of FFNN and LSTM with 100 and 50 neurons, respectively, with the dropout regularization method enabled the models to achieve highly accurate predictions on test sets with R^2 = 0.73 and 0.82 respectively. Finally, the ensemble of both architectures further improved prediction accuracy, achieving the highest R^2 value of 0.84. The SHAP model interpretation revealed that elevation, pH, NDVI, SI1, SI7 (CRSI), and OM exhibited significant positive and negative impacts on soil salinity prediction. Overall, deep learning models with SHAP interpretation proved their efficiency in accurate soil salinity predictions from various soil and remote sensing indicators.

Hence, this research clearly demonstrates the effectiveness of improved deep learning architectures as a valuable tool for accurate soil prediction. Moreover, integration of remote sensing and field collected soil sampled observations provides a more in-depth insight into the multiple factors of soil salinity, and demonstrates the potential of utilizing non-traditional,

efficient, and cost-effective methods for soil and agricultural resource management. Hence, this study underscores the significance of integrating multi-source data and advanced modeling techniques to address complex environmental problems and proposes sustainable land management practices to overcome the negative impacts of soil salinity on crop production.

References

[1] Zhu C, Ding J, Zhang Z, Wang J, Chen X, Han L, et al. Soil salinity dynamics in arid oases during irrigated and non-irrigated seasons. Land Degradation & Development. 2023;34(13):3823-35. https://doi.org/10.1002/ldr.4632

[2] Rajath E, Kalambukattu JG, Haldar D, Kumar S. Chapter 28 - Soil salinity mapping using multisensor data employing machine-learning technique: a case study from indo-gangetic plain. In: Dharumarajan S, Kaliraj S, Adhikari K, Lalitha M, Kumar N, editors. Remote Sensing of Soils: Elsevier; 2024. p. 439-53.

[3] Tedeschi A. Irrigated Agriculture on Saline Soils: A Perspective. Agronomy. 2020;10(11):1630.

[4] Wei Y, Ding J, Yang S, Wang F, Wang C. Soil salinity prediction based on scale-dependent relationships with environmental variables by discrete wavelet transform in the Tarim Basin. CATENA. 2021;196:104939. https://doi.org/10.1016/j.catena.2020.104939

[5] Zörb C, Geilfus CM, Dietz KJ. Salinity and crop yield. Plant Biology. 2019;21(S1):31-8. https://doi.org/10.1111/plb.12884

[6] Singh A. Soil salinity: A global threat to sustainable development. Soil Use and Management. 2022;38(1):39-67. https://doi.org/10.1111/sum.12772

[7] Singh A. Soil salinization management for sustainable development: A review. Journal of Environmental Management. 2021;277:111383. https://doi.org/10.1016/j.jenvman.2020.111383

[8] Eswar D, Karuppusamy R, Chellamuthu S. Drivers of soil salinity and their correlation with climate change. Current Opinion in Environmental Sustainability. 2021;50:310-8. https://doi.org/10.1016/j.cosust.2020.10.015

[9] Cuthbert RN, Weyl OLF, Wasserman RJ, Dick JTA, Froneman PW, Callaghan A, et al. Combined impacts of warming and salinisation on trophic interactions and mortality of a specialist ephemeral wetland predator. Freshwater Biology. 2019;64(9):1584-92. https://doi.org/10.1111/fwb.13353

[10] Measho S, Li F, Pellikka P, Tian C, Hirwa H, Xu N, et al. Soil Salinity Variations and Associated Implications for Agriculture and Land Resources Development Using Remote Sensing Datasets in Central Asia. Remote Sensing. 2022;14(10):2501.

[11] Paz AM, Castanheira N, Farzamian M, Paz MC, Gonçalves MC, Monteiro Santos FA, et al. Prediction of soil salinity and sodicity using electromagnetic conductivity imaging. Geoderma. 2020;361:114086. https://doi.org/10.1016/j.geoderma.2019.114086

[12] Zhang Z, Ding J, Zhu C, Wang J, Ma G, Ge X, et al. Strategies for the efficient estimation of soil organic matter in salt-affected soils through Vis-NIR spectroscopy: Optimal band combination algorithm and spectral degradation. Geoderma. 2021;382:114729. https://doi.org/10.1016/j.geoderma.2020.114729

[13] Singh P, Sharma S, Nisar S, Choudhary OP. Structural Stability and Organic Matter Stabilization in Soils: Differential Impacts of Soil Salinity and Sodicity. Journal of Soil Science and Plant Nutrition. 2023;23(2):1751-73. 10.1007/s42729-023-01136-3

[14] Corwin DL, Scudiero E. Field-scale apparent soil electrical conductivity. Soil Science Society of America Journal. 2020;84(5):1405-41. https://doi.org/10.1002/saj2.20153

[15] Singh A. Managing the salinization and drainage problems of irrigated areas through remote sensing and GIS techniques. Ecological Indicators. 2018;89:584-9. https://doi.org/10.1016/j.ecolind.2018.02.041

[16] Akça E, Aydin M, Kapur S, Kume T, Nagano T, Watanabe T, et al. Long-term monitoring of soil salinity in a semi-arid environment of Turkey. CATENA. 2020;193:104614. https://doi.org/10.1016/j.catena.2020.104614
[17] Wang F, Yang S, Wei Y, Shi Q, Ding J. Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: A case study in Tarim River Basin of southern Xinjiang, China. Science of The Total Environment. 2021;754:142030. https://doi.org/10.1016/j.scitotenv.2020.142030
[18] Ren D, Wei B, Xu X, Engel B, Li G, Huang Q, et al. Analyzing spatiotemporal characteristics of soil salinity in arid irrigated agro-ecosystems using integrated approaches. Geoderma. 2019;356:113935. https://doi.org/10.1016/j.geoderma.2019.113935
[19] Corwin DL, Scudiero E. Chapter One - Review of soil salinity assessment for agriculture across multiple scales using proximal and/or remote sensors. In: Sparks DL, editor. Advances in Agronomy: Academic Press; 2019. p. 1-130.
[20] Barbouchi M, Abdelfattah R, Chokmani K, Aissa NB, Lhissou R, Harti AE. Soil Salinity Characterization Using Polarimetric InSAR Coherence: Case Studies in Tunisia and Morocco. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015;8(8):3823-32. 10.1109/JSTARS.2014.2333535
[21] Vermeulen D, Van Niekerk A. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. Geoderma. 2017;299:1-12. https://doi.org/10.1016/j.geoderma.2017.03.013
[22] Avdan U, Kaplan G, Küçük Matcı D, Yiğit Avdan Z, Erdem F, Tuğba Mızık E, et al. Soil salinity prediction models constructed by different remote sensors. Physics and Chemistry of the Earth, Parts A/B/C. 2022;128:103230. https://doi.org/10.1016/j.pce.2022.103230
[23] El Harti A, Lhissou R, Chokmani K, Ouzemou J-e, Hassouna M, Bachaoui EM, et al. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices. International Journal of Applied Earth Observation and Geoinformation. 2016;50:64-73. https://doi.org/10.1016/j.jag.2016.03.008
[24] Scudiero E, Skaggs TH, Corwin DL. Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance. Remote Sensing of Environment. 2015;169:335-43. https://doi.org/10.1016/j.rse.2015.08.026
[25] Ge X, Ding J, Teng D, Xie B, Zhang X, Wang J, et al. Exploring the capability of Gaofen-5 hyperspectral data for assessing soil salinity risks. International Journal of Applied Earth Observation and Geoinformation. 2022;112:102969. https://doi.org/10.1016/j.jag.2022.102969
[26] Gorji T, Sertel E, Tanik A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. Ecological Indicators. 2017;74:384-91. https://doi.org/10.1016/j.ecolind.2016.11.043
[27] Mohammed S, Arshad S, Bashir B, Vad A, Alsalman A, Harsányi E. Machine learning driven forecasts of agricultural water quality from rainfall ionic characteristics in Central Europe. Agricultural Water Management. 2024;293:108690. https://doi.org/10.1016/j.agwat.2024.108690
[28] Jamei M, Ali M, Karbasi M, Karimi B, Jahannemaei N, Farooque AA, et al. Monthly sodium adsorption ratio forecasting in rivers using a dual interpretable glass-box complementary intelligent system: Hybridization of ensemble TVF-EMD-VMD, Boruta-SHAP, and eXplainable GPR. Expert Systems with Applications. 2024;237:121512. https://doi.org/10.1016/j.eswa.2023.121512
[29] Xiao C, Ji Q, Chen J, Zhang F, Li Y, Fan J, et al. Prediction of soil salinity parameters using machine learning models in an arid region of northwest China. Computers and Electronics in Agriculture. 2023;204:107512. https://doi.org/10.1016/j.compag.2022.107512
[30] Sarkar SK, Rudra RR, Sohan AR, Das PC, Ekram KMM, Talukdar S, et al. Coupling of machine learning and remote sensing for soil salinity mapping in coastal area of Bangladesh. Scientific Reports. 2023;13(1):17056. 10.1038/s41598-023-44132-4
[31] Zhao S, Ayoubi S, Mousavi SR, Mireei SA, Shahpouri F, Wu S-x, et al. Integrating proximal soil sensing data and environmental variables to enhance the prediction accuracy for soil salinity and sodicity in a region of Xinjiang Province, China. Journal of Environmental Management. 2024;364:121311. https://doi.org/10.1016/j.jenvman.2024.121311

[32] Bandak S, Movahedi-Naeini SA, Mehri S, Lotfata A. A longitudinal analysis of soil salinity changes using remotely sensed imageries. Scientific Reports. 2024;14(1):10383. 10.1038/s41598-024-60033-6
[33] Zhou M, Li Y. Digital Mapping and Scenario Prediction of Soil Salinity in Coastal Lands Based on Multi-Source Data Combined with Machine Learning Algorithms. Remote Sensing. 2024;16(14):2681.
[34] Wang L, Hu P, Zheng H, Liu Y, Cao X, Hellwich O, et al. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images. Geoderma. 2023;430:116321. https://doi.org/10.1016/j.geoderma.2022.116321
[35] Andrade Foronda D, Colinet G. Prediction of Soil Salinity/Sodicity and Salt-Affected Soil Classes from Soluble Salt Ions Using Machine Learning Algorithms. Soil Systems. 2023;7(2):47.
[36] Kaplan G, Gašparović M, Alqasemi AS, Aldhaheri A, Abuelgasim A, Ibrahim M. Soil salinity prediction using Machine Learning and Sentinel – 2 Remote Sensing Data in Hyper – Arid areas. Physics and Chemistry of the Earth, Parts A/B/C. 2023;130:103400. https://doi.org/10.1016/j.pce.2023.103400
[37] Ge X, Ding J, Teng D, Wang J, Huo T, Jin X, et al. Updated soil salinity with fine spatial resolution and high accuracy: The synergy of Sentinel-2 MSI, environmental covariates and hybrid machine learning approaches. CATENA. 2022;212:106054. https://doi.org/10.1016/j.catena.2022.106054
[38] Wang F, Shi Z, Biswas A, Yang S, Ding J. Multi-algorithm comparison for predicting soil salinity. Geoderma. 2020;365:114211. https://doi.org/10.1016/j.geoderma.2020.114211
[39] Nawaz A, Farooq M, Ul-Allah S, Gogoi N, Lal R, Siddique KHM. Sustainable Soil Management for Food Security in South Asia. Journal of Soil Science and Plant Nutrition. 2021;21(1):258-75. 10.1007/s42729-020-00358-z
[40] Qureshi AS, McCornick PG, Qadir M, Aslam Z. Managing salinity and waterlogging in the Indus Basin of Pakistan. Agricultural Water Management. 2008;95(1):1-10. https://doi.org/10.1016/j.agwat.2007.09.014
[41] Ashraf M, Fatima B, Hasan Fu, Salam HA. Adapting to Salinity in the Southern Indus Basin Project: Policy Review. Islamabad: Pakistan Council of Research in Water Resources (PCRWR); 2022. p. 52.
[42] Syed A, Sarwar G, Shah SH, Muhammad S. Soil Salinity Research in 21st Century in Pakistan: Its Impact on Availability of Plant Nutrients, Growth and Yield of Crops. Communications in Soil Science and Plant Analysis. 2021;52(3):183-200. 10.1080/00103624.2020.1854294
[43] Haq Yu, Shahbaz M, Asif HMS, Al-Laith A, Alsabban WH. Spatial Mapping of Soil Salinity Using Machine Learning and Remote Sensing in Kot Addu, Pakistan. Sustainability. 2023;15(17):12943.
[44] Ijaz M, Ahmad HR, Bibi S, Ayub MA, Khalid S. Soil salinity detection and monitoring using Landsat data: a case study from Kot Addu, Pakistan. Arabian Journal of Geosciences. 2020;13(13):510. 10.1007/s12517-020-05572-8
[45] Arshad S, Kazmi JH, Shaikh S, Fatima M, Faheem Z, Asif M, et al. Geospatial assessment of early summer heatwaves, droughts, and their relationship with vegetation and soil moisture in the arid region of Southern Punjab, Pakistan. Journal of Water and Climate Change. 2022;13(11):4105-29.
[46] Tounkara F, Ehsan M, Nasar Iqbal M, Al-Ansari N, Hajana MI, Shafi A, et al. Analyzing the seismic attributes, structural and petrophysical analyses of the Lower Goru Formation: A case study from Middle Indus Basin Pakistan. Frontiers in Earth Science. 2023;10. 10.3389/feart.2022.1034874
[47] Ali F, Zhang S, Abbas S, Hanif M, Zhang Y, Mohibullah M. Discovery of the Lower-Middle Jurassic high potential source rocks based on the geochemical investigation from the Indus Basin, Pakistan. Journal of Petroleum Science and Engineering. 2022;208:109374. https://doi.org/10.1016/j.petrol.2021.109374
[48] Arshad S, Kazmi JH, Prodhan FA, Mohammed S. Exploring dynamic response of agrometeorological droughts towards winter wheat yield loss risk using machine learning approach at a regional scale in Pakistan. Field Crops Research. 2023;302:109057. https://doi.org/10.1016/j.fcr.2023.109057
[49] FAO. Soil Fertility Atlas of Pakistan: The Punjab Province. In: Ahmad W, Niino Y, Zia MH, Mahmood K, Ahmad N, Salim M, et al., editors.: Islamabad, Pakistan; 2017.

[50] Ashraf M, Fatima B, Hasan Fu, Salam HA. Adapting to Salinity in the Southern Indus Basin (ASSIB) Project. Islamabd: Pakistan Council of Research in Water Resources (PCRWR); 2022.
[51] Iqbal N, Ashraf M, Imran M, Salam HA, Hasan Fu. Groundwater Investigations and Mapping in the Lower Indus plain. Islamabad: Pakistan Council of Research in Water Resources (PCRWR); 2020. p. 70.
[52] Ali Q, Shabaan M, Ashraf S, Kamran M, Zulfiqar U, Ahmad M, et al. Comparative efficacy of different salt tolerant rhizobial inoculants in improving growth and productivity of Vigna radiata L. under salt stress. Scientific Reports. 2023;13(1):17442. 10.1038/s41598-023-44433-8
[53] Scudiero E, Skaggs TH, Corwin DL. Comparative regional-scale soil salinity assessment with near-ground apparent electrical conductivity and remote sensing canopy reflectance. Ecological Indicators. 2016;70:276-84. https://doi.org/10.1016/j.ecolind.2016.06.015
[54] Lobell DB, Lesch SM, Corwin DL, Ulmer MG, Anderson KA, Potts DJ, et al. Regional-scale Assessment of Soil Salinity in the Red River Valley Using Multi-year MODIS EVI and NDVI. Journal of Environmental Quality. 2010;39(1):35-41. https://doi.org/10.2134/jeq2009.0140
[55] Zeyliger AM, Muzalevskiy KV, Zinchenko EV, Ermolaeva OS. Field test of the surface soil moisture mapping using Sentinel-1 radar data. Science of The Total Environment. 2022;807:151121. https://doi.org/10.1016/j.scitotenv.2021.151121
[56] Azizi K, Garosi Y, Ayoubi S, Tajik S. Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. Soil and Tillage Research. 2023;229:105681. https://doi.org/10.1016/j.still.2023.105681
[57] Hoa PV, Giang NV, Binh NA, Hai LVH, Pham T-D, Hasanlou M, et al. Soil Salinity Mapping Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms: A Case Study at Ben Tre Province of the Mekong River Delta (Vietnam). Remote Sensing. 2019;11(2):128.
[58] Peng J, Biswas A, Jiang Q, Zhao R, Hu J, Hu B, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. Geoderma. 2019;337:1309-19. https://doi.org/10.1016/j.geoderma.2018.08.006
[59] Buchanan BP, Fleming M, Schneider RL, Richards BK, Archibald J, Qiu Z, et al. Evaluating topographic wetness indices across central New York agricultural landscapes. Hydrol Earth Syst Sci. 2014;18(8):3279-99. 10.5194/hess-18-3279-2014
[60] Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 1979;8(2):127-50. https://doi.org/10.1016/0034-4257(79)90013-0
[61] Jiang Z, Huete AR, Didan K, Miura T. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment. 2008;112(10):3833-45. https://doi.org/10.1016/j.rse.2008.06.006
[62] Stafford JV, Vincini M, Frazzi E, D'Alessio P. Precision agriculture '07: Wageningen Academic, 2007.
[63] Khan NM, Rastoskuev VV, Sato Y, Shiozawa S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. Agricultural Water Management. 2005;77(1):96-109. https://doi.org/10.1016/j.agwat.2004.09.038
[64] Douaoui AEK, Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. Geoderma. 2006;134(1):217-30. https://doi.org/10.1016/j.geoderma.2005.10.009
[65] Wang J, Ding J, Yu D, Ma X, Zhang Z, Ge X, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. Geoderma. 2019;353:172-87. https://doi.org/10.1016/j.geoderma.2019.06.040
[66] Scudiero E, Skaggs TH, Corwin DL. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. Geoderma Regional. 2014;2-3:82-90. https://doi.org/10.1016/j.geodrs.2014.10.004
[67] Ma G, Ding J, Han L, Zhang Z, Ran S. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. Regional Sustainability. 2021;2(2):177-88. https://doi.org/10.1016/j.regsus.2021.06.001
[68] Kozak M, Wnuk A. Including the Tukey Mean-Difference (Bland–Altman) Plot in a Statistics Course. Teaching Statistics. 2014;36(3):83-7. https://doi.org/10.1111/test.12032
[69] Greenacre M, Groenen PJF, Hastie T, D'Enza AI, Markos A, Tuzhilina E. Principal component analysis. Nature Reviews Methods Primers. 2022;2(1):100. 10.1038/s43586-022-00184-w

[70] Taghizadeh-Mehrjardi R, Schmidt K, Toomanian N, Heung B, Behrens T, Mosavi A, et al. Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. Geoderma. 2021;383:114793. https://doi.org/10.1016/j.geoderma.2020.114793
[71] Tao H, Abba SI, Al-Areeq AM, Tangang F, Samantaray S, Sahoo A, et al. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions. Engineering Applications of Artificial Intelligence. 2024;129:107559. https://doi.org/10.1016/j.engappai.2023.107559
[72] Chen Q, Meng Z, Liu X, Jin Q, Su R. Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE. Genes. 2018;9(6):301.
[73] Pullanagari RR, Kereszturi G, Yule I. Integrating Airborne Hyperspectral, Topographic, and Soil Data for Estimating Pasture Quality Using Recursive Feature Elimination with Random Forest Regression. Remote Sensing. 2018;10(7):1117.
[74] Wang Y, Li Y. Mapping the ratoon rice suitability region in China using random forest and recursive feature elimination modeling. Field Crops Research. 2023;301:109016. https://doi.org/10.1016/j.fcr.2023.109016
[75] Ma L, Khorasani K. A new strategy for adaptively constructing multilayer feedforward neural networks. Neurocomputing. 2003;51:361-85. https://doi.org/10.1016/S0925-2312(02)00597-0
[76] Hemeida AM, Hassan SA, Mohamed A-AA, Alkhalaf S, Mahmoud MM, Senjyu T, et al. Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research. Ain Shams Engineering Journal. 2020;11(3):659-75. https://doi.org/10.1016/j.asej.2020.01.007
[77] Ahlawat S. Recurrent Neural Networks. In: Ahlawat S, editor. Reinforcement Learning for Finance: Solve Problems in Finance with CNN and RNN Using the TensorFlow Library. Berkeley, CA: Apress; 2023. p. 177-232.
[78] Naguib M, Kollmeyer P, Vidal C, Emadi A. Accurate Surface Temperature Estimation of Lithium-Ion Batteries Using Feedforward and Recurrent Artificial Neural Networks. Conference Accurate Surface Temperature Estimation of Lithium-Ion Batteries Using Feedforward and Recurrent Artificial Neural Networks. p. 52-7.
[79] Zakhrouf M, Hamid B, Kim S, Madani S. Novel insights for streamflow forecasting based on deep learning models combined the evolutionary optimization algorithm. Physical Geography. 2023;44(1):31-54. 10.1080/02723646.2021.1943126
[80] Kim Soon G, Kim On C, Mohd Rusli N, Soo Fun T, Alfred R, Tse Guan T. Comparison of simple feedforward neural network, recurrent neural network and ensemble neural networks in phishing detection. Journal of Physics: Conference Series. 2020;1502(1):012033. 10.1088/1742-6596/1502/1/012033
[81] Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32. 10.1023/A:1010933404324
[82] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: Association for Computing Machinery; 2016. p. 1135–44.
[83] Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering. 2018;2(10):749-60. 10.1038/s41551-018-0304-0
[84] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems. 2014;41(3):647-65. 10.1007/s10115-013-0679-x
[85] Zhong L, Guo X, Ding M, Ye Y, Jiang Y, Zhu Q, et al. SHAP values accurately explain the difference in modeling accuracy of convolution neural network between soil full-spectrum and feature-spectrum. Computers and Electronics in Agriculture. 2024;217:108627. https://doi.org/10.1016/j.compag.2024.108627
[86] Stavi I, Thevs N, Priori S. Soil Salinity and Sodicity in Drylands: A Review of Causes, Effects, Monitoring, and Restoration Measures. Frontiers in Environmental Science. 2021;9. 10.3389/fenvs.2021.712831
[87] Corwin DL. Chapter Two - Dealing with the impact of climate change-induced drought on the management of soil salinity under irrigated agriculture. In: Sparks DL, editor. Advances in Agronomy: Academic Press; 2024. p. 67-124.

[88] Shahid SA, Zaman M, Heng L. Soil Salinity: Historical Perspectives and a World Overview of the Problem. In: Zaman M, Shahid SA, Heng L, editors. Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques. Cham: Springer International Publishing; 2018. p. 43-53.
[89] Condom N, Kuper M, Marlet S, Valles V, Kijne J. Salinization, alkalinization and sodification in Punjab (Pakistan): characterization of the geochemical and physical processes of degradation. Land Degradation & Development. 1999;10(2):123-40. https://doi.org/10.1002/(SICI)1099-145X(199903/04)10:2<123::AID-LDR321>3.0.CO;2-V
[90] Cox C, Jin L, Ganjegunte G, Borrok D, Lougheed V, Ma L. Soil quality changes due to flood irrigation in agricultural fields along the Rio Grande in western Texas. Applied Geochemistry. 2018;90:87-100. https://doi.org/10.1016/j.apgeochem.2017.12.019
[91] Davidson AP. Soil salinity, a major constraint to irrigated agriculture in the Punjab region of Pakistan: Contributing factors and strategies for amelioration. American Journal of Alternative Agriculture. 2000;15(4):154-9. 10.1017/S0889189300008729
[92] Qureshi A, Asghar M, Ahmad S, Masih I. Sustaining crop production under saline groundwater conditions: A case study from Pakistan. Australian Journal of Agricultural Sciences. 2004;54(2):421-31.
[93] Latif M, Ahmad MZ. Groundwater and soil salinity variations in a canal command area in Pakistan. Irrigation and Drainage. 2009;58(4):456-68. https://doi.org/10.1002/ird.417
[94] Qureshi R, Barrett-Lennard E. Saline agriculture for irrigated land in Pakistan: a handbook. Canberra: Australian Centre for International Agricultural Research (ACIAR), 1998.
[95] Quereshi AS, Sarwar A. Managing salinity in the Indus Basin of Pakistan. International Journal of River Basin Management. 2009;7(2):111-7. 10.1080/15715124.2009.9635373
[96] Bashir MA, Rehim A, Liu J, Imran M, Liu H, Suleman M, et al. Soil survey techniques determine nutrient status in soil profile and metal retention by calcium carbonate. CATENA. 2019;173:141-9. https://doi.org/10.1016/j.catena.2018.10.015
[97] Habibi V, Ahmadi H, Jafari M, Moeini A. Mapping soil salinity using a combined spectral and topographical indices with artificial neural network. PLOS ONE. 2021;16(5):e0228494. 10.1371/journal.pone.0228494
[98] Shahrayini E, Noroozi AA. Modeling and Mapping of Soil Salinity and Alkalinity Using Remote Sensing Data and Topographic Factors: a Case Study in Iran. Environmental Modeling & Assessment. 2022;27(5):901-13. 10.1007/s10666-022-09823-8
[99] Yang F, An F, Ma H, Wang Z, Zhou X, Liu Z. Variations on Soil Salinity and Sodicity and Its Driving Factors Analysis under Microtopography in Different Hydrological Conditions. Water. 2016;8(6):227.
[100] Abd El-Hamid HT, Alshehri F, El-Zeiny AM, Nour-Eldin H. Remote sensing and statistical analyses for exploration and prediction of soil salinity in a vulnerable area to seawater intrusion. Marine Pollution Bulletin. 2023;187:114555. https://doi.org/10.1016/j.marpolbul.2022.114555
[101] Guo B, Yang X, Yang M, Sun D, Zhu W, Zhu D, et al. Mapping soil salinity using a combination of vegetation index time series and single-temporal remote sensing images in the Yellow River Delta, China. CATENA. 2023;231:107313. https://doi.org/10.1016/j.catena.2023.107313
[102] Pérez González ME, García Rodríguez MdP, González-Quiñones V, Jiménez Ballesta R. Spatial variability of soil quality in the surroundings of a saline lake environment. Environmental Geology. 2006;51(1):143-9. 10.1007/s00254-006-0317-y
[103] Sarkar SK, Rudra RR, Nur MS, Das PC. Partial least-squares regression for soil salinity mapping in Bangladesh. Ecological Indicators. 2023;154:110825. https://doi.org/10.1016/j.ecolind.2023.110825
[104] Liu J, Yang K, Tariq A, Lu L, Soufan W, El Sabagh A. Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods. The Egyptian Journal of Remote Sensing and Space Sciences. 2023;26(3):415-26. https://doi.org/10.1016/j.ejrs.2023.05.005
[105] Zhang Q, Li L, Sun R, Zhu D, Zhang C, Chen Q. Retrieval of the Soil Salinity From Sentinel-1 Dual-Polarized SAR Data Based on Deep Neural Network Regression. IEEE Geoscience and Remote Sensing Letters. 2022;19:1-5. 10.1109/LGRS.2020.3041059

[106] Tripathi A, Tiwari RK. A simplified subsurface soil salinity estimation using synergy of SENTINEL-1 SAR and SENTINEL-2 multispectral satellite data, for early stages of wheat crop growth in Rupnagar, Punjab, India. Land Degradation & Development. 2021;32(14):3905-19. https://doi.org/10.1002/ldr.4009
[107] Siqueira RG, Moquedace CM, Fernandes-Filho EI, Schaefer CEGR, Francelino MR, Sacramento IF, et al. Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica. CATENA. 2024;235:107677. https://doi.org/10.1016/j.catena.2023.107677
[108] Tarolli P, Luo J, Park E, Barcaccia G, Masin R. Soil salinization in agriculture: Mitigation and adaptation strategies combining nature-based solutions and bioengineering. iScience. 2024;27(2). 10.1016/j.isci.2024.108830
[109] Mosaid H, Barakat A, John K, Faouzi E, Bustillo V, El Garnaoui M, et al. Improved soil carbon stock spatial prediction in a Mediterranean soil erosion site through robust machine learning techniques. Environmental Monitoring and Assessment. 2024;196(2):130. 10.1007/s10661-024-12294-x
[110] Raja SP, Sawicka B, Stamenkovic Z, Mariammal G. Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. IEEE Access. 2022;10:23625-41. 10.1109/ACCESS.2022.3154350
[111] Boudibi S, Sakaa B, Benguega Z, Fadlaoui H, Othman T, Bouzidi N. Spatial prediction and modeling of soil salinity using simple cokriging, artificial neural networks, and support vector machines in El Outaya plain, Biskra, southeastern Algeria. Acta Geochimica. 2021;40(3):390-408. 10.1007/s11631-020-00444-0
[112] Pouladi N, Jafarzadeh AA, Shahbazi F, Ghorbani MA. Design and implementation of a hybrid MLP-FFA model for soil salinity prediction. Environmental Earth Sciences. 2019;78(5):159. 10.1007/s12665-019-8159-6
[113] Barzegar R, Asghari Moghaddam A. Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. Modeling Earth Systems and Environment. 2016;2(1):26. 10.1007/s40808-015-0072-8
[114] Park S-H, Lee B-Y, Kim M-J, Sang W, Seo MC, Baek J-K, et al. Development of a Soil Moisture Prediction Model Based on Recurrent Neural Network Long Short-Term Memory (RNN-LSTM) in Soybean Cultivation. Sensors. 2023;23(4):1976.
[115] Mohammadifar A, Gholami H, Golzari S. Assessment of the uncertainty and interpretability of deep learning models for mapping soil salinity using DeepQuantreg and game theory. Scientific Reports. 2022;12(1):15167. 10.1038/s41598-022-19357-4
[116] Zaki MT, Abdul-Aziz OI. Predicting greenhouse gas fluxes in coastal salt marshes using artificial neural networks. Wetlands. 2022;42(5):37. 10.1007/s13157-022-01558-2
[117] Abba SI, Benaafi M, Usman AG, Aljundi IH. Inverse groundwater salinization modeling in a sandstone's aquifer using stand-alone models with an improved non-linear ensemble machine learning technique. Journal of King Saud University - Computer and Information Sciences. 2022;34(10, Part A):8162-75. https://doi.org/10.1016/j.jksuci.2022.08.002
[118] Nandini GS, Kumar APS, K C. Dropout technique for image classification based on extreme learning machine. Global Transitions Proceedings. 2021;2(1):111-6. https://doi.org/10.1016/j.gltp.2021.01.015
[119] Chen Y, Song L, Liu Y, Yang L, Li D. A Review of the Artificial Neural Network Models for Water Quality Prediction. Applied Sciences. 2020;10(17):5776.
[120] Khan J, Lee E, Kim K. A higher prediction accuracy–based alpha–beta filter algorithm using the feedforward artificial neural network. CAAI Transactions on Intelligence Technology. 2023;8(4):1124-39. https://doi.org/10.1049/cit2.12148
[121] O'Donncha F, Hu Y, Palmes P, Burke M, Filgueira R, Grant J. A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales. Ecological Informatics. 2022;69:101687. https://doi.org/10.1016/j.ecoinf.2022.101687
[122] Tian H, Wang P, Tansey K, Zhang J, Zhang S, Li H. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. Agricultural and Forest Meteorology. 2021;310:108629. https://doi.org/10.1016/j.agrformet.2021.108629

[123] Mohammed S, Arshad S, Bashir B, Ata B, Al-Dalahmeh M, Alsalman A, et al. Evaluating machine learning performance in predicting sodium adsorption ratio for sustainable soil-water management in the eastern Mediterranean. Journal of Environmental Management. 2024;370:122640. https://doi.org/10.1016/j.jenvman.2024.122640
[124] Solangi GS, Siyal A, Babar M, Siyal P. Spatial analysis of soil salinity in the Indus River Delta, Pakistan. Engineering, Technology & Applied Science Research. 2019;9(3):4271-5. https://doi.org/10.48084/etasr.2818
[125] Wang S, Chen Y, Wang M, Li J. Performance Comparison of Machine Learning Algorithms for Estimating the Soil Salinity of Salt-Affected Soil Using Field Spectral Data. Remote Sensing. 2019;11(22):2605.
[126] Bannari A, Guedon AM, El-Harti A, Cherkaoui FZ, El-Ghmari A. Characterization of Slightly and Moderately Saline and Sodic Soils in Irrigated Agricultural Land using Simulated Data of Advanced Land Imaging (EO-1) Sensor. Communications in Soil Science and Plant Analysis. 2008;39(19-20):2795-811. 10.1080/00103620802432717
[127] Tian F, Hou M, Qiu Y, Zhang T, Yuan Y. Salinity stress effects on transpiration and plant growth under different salinity soil levels based on thermal infrared remote (TIR) technique. Geoderma. 2020;357:113961. https://doi.org/10.1016/j.geoderma.2019.113961

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Appendix Table 1. Descriptive statistics of all variables (soil and remote sensing)

	texture	Mean	Median	SD	Range	Minimum	Maximum	Skewness	Kurtosis
EC	Loam	3.589	2.54	2.8265	10.93	0.57	11.5	1.1292	0.0741
	Sandy Loam	8.518	9.39	4.4225	14.24	1.26	15.5	-0.2756	-0.7604
pH	Loam	8.34	8.3	0.4605	3.79	7.4	11.19	2.9422	16.5301

	Sandy Loam	9.009	8.89	0.9442	3.5	7.8	11.3	0.8178	-0.0421
OM	Loam	0.712	0.66	0.3365	1.74	0.17	1.91	1.6759	3.2665
	Sandy Loam	0.51	0.49	0.2324	1.18	0.31	1.49	3.6325	15.532
P (2)	Loam	7.102	6.85	4.1508	21.8	1.5	23.3	1.2328	2.259
	Sandy Loam	5.58	5.5	2.0034	8.2	3	11.2	1.1265	1.5185
K (2)	Loam	129.221	110	65.3442	326	15	341	1.2477	1.561
	Sandy Loam	110.87	104	50.5554	274	57	331	3.9863	18.1609
SP	Loam	35.663	36	4.639	38	8	46	-2.063	14.2572
	Sandy Loam	33.739	36	4.8544	17	24	41	-0.3455	-1.0539
CVI	Loam	1.989	2.032	0.2475	1.169	1.2313	2.4	-0.8078	0.3842
	Sandy Loam	1.837	1.925	0.348	1.313	1.0714	2.385	-0.9301	0.3878
EVI	Loam	0.324	0.329	0.1225	0.461	0.0522	0.513	-0.6132	-0.3835
	Sandy Loam	0.257	0.27	0.1457	0.554	0.0128	0.567	0.186	-0.5001
NDVI	Loam	0.281	0.299	0.1102	0.415	0.0379	0.453	-0.6382	-0.3944
	Sandy Loam	0.216	0.191	0.1182	0.427	0.0493	0.477	0.5537	-0.5249
SI1	Loam	0.209	0.193	0.0515	0.218	0.1344	0.352	1.1958	0.7279
	Sandy Loam	0.231	0.229	0.0474	0.149	0.1505	0.299	-0.2396	-1.3333
SI2	Loam	0.228	0.212	0.0562	0.254	0.1472	0.402	1.4537	1.7115
	Sandy Loam	0.253	0.251	0.043	0.153	0.1831	0.336	-0.0741	-1.0147
SI3	Loam	0.321	0.3	0.08	0.363	0.2083	0.571	1.5052	1.8944
	Sandy Loam	0.365	0.352	0.0877	0.33	0.2437	0.574	0.6774	-0.2122
SI4	Loam	0.472	0.464	0.0504	0.285	0.3566	0.641	1.2254	2.3967
	Sandy Loam	0.487	0.459	0.0656	0.252	0.3984	0.651	1.5236	1.7368
SI5	Loam	0.52	0.504	0.0679	0.391	0.3497	0.741	1.1976	2.3165
	Sandy Loam	0.532	0.507	0.0646	0.242	0.4504	0.692	1.3772	1.3383
SI6	Loam	0.528	0.512	0.0647	0.346	0.3978	0.743	1.4184	2.3733
	Sandy Loam	0.541	0.517	0.0628	0.243	0.4588	0.701	1.4319	1.576
SI7	Loam	0.647	0.658	0.0469	0.294	0.4105	0.704	-2.0685	6.914
	Sandy Loam	0.664	0.665	0.0672	0.356	0.5182	0.875	1.0819	4.2189
SARVV	Loam	-10.812	-10.28	3.2504	16.256	-20.0055	-3.749	-0.8287	0.6702
	Sandy Loam	-11.731	-10.728	3.6013	13.613	-19.5306	-5.918	-0.7731	-0.1123
SARVH	Loam	-17.63	-16.185	3.6622	15.27	-29.0952	-13.825	-1.5534	1.4748

	Sandy Loam	-18.907	-17.715	3.7868	14.944	-29.4491	-14.505	-1.1333	1.1713
ASPECT	Loam	164.715	180	109.0071	338.952	0	338.952	-0.1572	-1.3872
	Sandy Loam	172.757	180	105.9599	330.055	0	330.055	-0.2166	-1.0586
ELEVATION	Loam	111.105	100	31.0904	92	73	165	0.463	-1.385
	Sandy Loam	143.957	156	28.8798	99	77	176	-1.2508	0.3041
SLOPE	Loam	2.065	1.418	1.2883	5.931	0	5.931	1.2955	1.2115
	Sandy Loam	1.672	1.418	1.0544	3.504	0	3.504	0.2037	-0.652
TWI	Loam	-11.488	-15.453	7.3464	20.979	-17.8835	3.096	1.0175	-0.8352
	Sandy Loam	-12.911	-15.802	6.0061	16.807	-17.4248	-0.618	1.3736	0.2654

Appendix Table 2. Principal Components with eigenvalue and % of variance explained

Principal Component Number	Eigenvalue	Percentage of Variance (%)	Cumulative (%)
1	7.38396	33.56343	33.56343
2	2.41271	10.96686	44.5303
3	1.64104	7.45928	51.98958
4	1.34847	6.12939	58.11898
5	1.23319	5.60543	63.7244
6	1.08307	4.92307	68.64747
7	0.95764	4.35291	73.00037
8	0.89049	4.0477	77.04807
9	0.76803	3.49105	80.53912
10	0.7261	3.30047	83.83959
11	0.56848	2.58401	86.4236
12	0.54061	2.45731	88.88091
13	0.46487	2.11304	90.99395
14	0.43353	1.97057	92.96452
15	0.36441	1.6564	94.62092
16	0.29016	1.3189	95.93982
17	0.25962	1.18011	97.11993
18	0.21796	0.99071	98.11065
19	0.14078	0.63989	98.75053
20	0.12344	0.56108	99.31162
21	0.09089	0.41312	99.72474
22	0.06056	0.27526	100

Appendix Table 3. Coefficients/ loadings of variables in each PC

var	PC1	PC2	PC3
EC	**0.2936**	-0.2248	0.2274

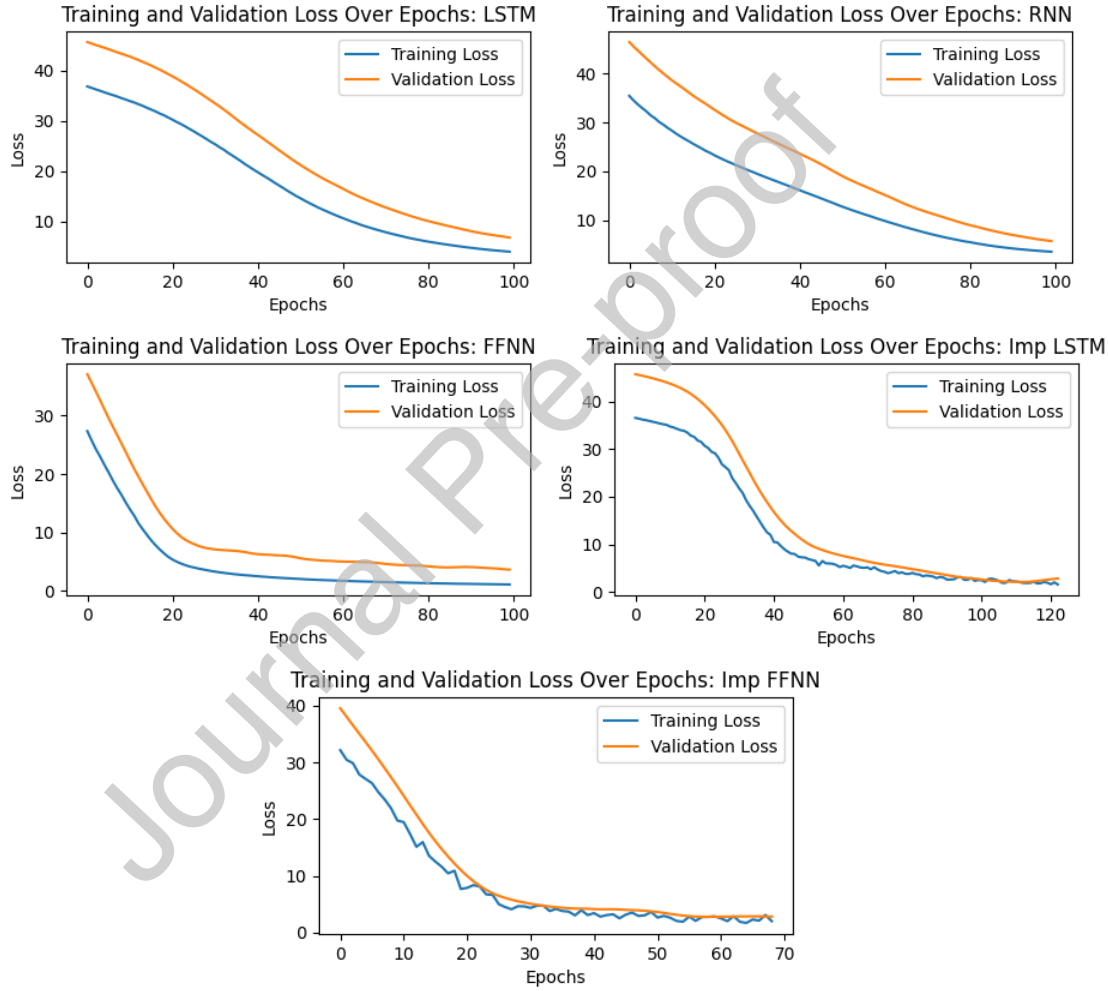
pH	**0.176**	-0.31902	0.21526
OM	-0.11777	**0.39258**	0.28296
P	-0.0821	**0.32972**	0.13982
K	-0.08155	**0.2955**	0.23498
Sp	-0.0783	0.05051	**0.36289**
Aspect	0.06206	0.12886	**0.27842**
Elevation	**0.24281**	-0.37498	0.0369
Slope	0.02198	**0.32225**	-0.17313
CVI	-0.22623	-0.08089	-0.22228
EVI	-0.30141	-0.03344	-0.07486
NDVI	-0.2924	-0.00887	-0.10693
SI1	**0.25327**	-0.08668	0.02147
SI2	**0.28297**	0.08024	-0.11428
SI3	**0.31938**	0.14843	-0.03501
SI4	**0.25137**	0.14371	0.09571
SI5	**0.26533**	0.28832	0.02372
SI6	**0.27342**	0.23476	-0.06943
SI7	0.05435	-0.0825	**0.4601**
SARVV	**-0.23597**	-0.0845	**0.35178**
SARVH	**-0.20529**	-0.12635	**0.27266**
TWI	-0.05713	-0.03742	-0.032

Appendix Table 4. Evaluation metrics of DL models for soil EC prediction

	Train set						**Test set**					
Model	**MAE**	**MSE**	**RMSE**	**R^2**	**MAPE**	**NSE**	**MAE**	**MSE**	**RMSE**	**R^2**	**MAPE**	**NSE**
FFNN	0.9	1.66	1.289	**0.889**	30.4	0.889	1.33	3.66	1.915	0.694	43.2	0.694
RNN	1.613	3.84	1.961	0.742	51.8	0.742	1.613	3.84	1.961	0.742	51.8	0.742
LSTM	1.698	4.34	2.083	0.709	52.4	0.709	1.545	3.44	1.855	0.713	57.4	0.713
imp-FFNN	1.109	1.94	1.394	**0.87**	40	0.87	1.231	3.22	1.796	**0.731**	41	0.731
imp-LSTM	1.032	1.74	1.322	**0.883**	35	0.883	1.036	2.11	1.453	**0.824**	33.9	**0.824**
Ensemble	1.02	1.7	1.306	**0.886**	36.3	0.886	1.014	1.91	1.382	**0.841**	35	**0.841**

Appendix Table 5. Model's generalization assessment based on training and validation loss

	Final Training loss	Final Validation loss	Minimum Validation loss	Mean Training loss	Mean Validation loss
FFNN	**1.095**	**2.596**	**2.596**	**8.896**	**11.292**
RNN	3.871	5.683	5.683	15.311	20.822
LSTM	3.837	7.156	7.156	16.579	23.521
imp-FFNN	**2.329**	**3.671**	**3.232**	**10.090**	**12.441**
imp-LSTM	**1.563**	**2.702**	**2.232**	**11.202**	**14.731**



Appendix Fig 1. Model's generalization from training and validation loss over the epochs for each DL model