DEEPSALT: Bridging Laboratory and Satellite Spectra through Domain Adaptation and Knowledge Distillation for Large-Scale Soil Salinity Estimation

Rupasree Dey[1], Abdul Matin[1], Everett Lewark[1], Tanjim Bin Faruk[1], Andrei Bachinin[1],
Sam Leuthold[2], M. Francesca Cotrufo[2], Shrideep Pallickara[1], and Sangmi Lee Pallickara[1]

[1]Department of Computer Science, Colorado State University, USA
[2]Department of Soil and Crop Sciences, Colorado State University, USA
{rupasree.dey, abdul.matin, elewark, tanjim.faruk, andrei.bachinin,
sam.leuthold, francesca.cotrufo, shrideep.pallickara, sangmi.pallickara}@colostate.edu

Abstract—Soil salinization poses a significant threat to both ecosystems and agriculture because it limits plants' ability to absorb water and, in doing so, reduces crop productivity. This phenomenon alters the soil's spectral properties, creating a measurable relationship between salinity and light reflectance that enables remote monitoring. While laboratory spectroscopy provides precise measurements, its reliance on in-situ sampling limits scalability to regional or global levels. Conversely, hyperspectral satellite imagery enables wide-area observation but lacks the fine-grained interpretability of laboratory instruments. To bridge this gap, we introduce DEEPSALT, a deep-learning-based spectral transfer framework that leverages knowledge distillation and a novel *Spectral Adaptation Unit* to transfer high-resolution spectral insights from laboratory-based spectroscopy to satellite-based hyperspectral sensing. Our approach eliminates the need for extensive ground sampling while enabling accurate, large-scale salinity estimation, as demonstrated through comprehensive empirical benchmarks. DEEPSALT achieves significant performance gains over methods without explicit domain adaptation, underscoring the impact of the proposed *Spectral Adaptation Unit* and the knowledge distillation strategy. The model also effectively generalized to unseen geographic regions, explaining a substantial portion of the salinity variance.

Index Terms—Transformer, Knowledge Distillation, Domain Adaptation, Hyperspectral Imagery, Soil Salinity Estimation

I. INTRODUCTION

Soil salinization poses a significant threat to agriculture and ecosystems. The accumulation of salt creates osmotic stress [1], which reduces the uptake of plant water and nutrients, reduces crop yields, and accelerates land degradation. These impacts are most pronounced in arid and semi-arid regions, where salinity hotspots directly undermine food security.

Remote sensing offers an efficient pathway for monitoring salinity at large scales. Compared with costly and sparse field sampling, satellite observations provide wall-to-wall coverage and a measurable link between spectral reflectance and soil salinity. Among these, hyperspectral data [2] have emerged as a powerful tool to map salinity dynamics due to their ability to capture subtle spectral features. However, the salinity estimation poses unique challenges. Unlike vegetation (with strong red-edge signals) or water (with distinct absorption features), salts such as NaCl produce weak responses in satellite-accessible ranges [3]. Salinity distributions obtained from remote sensing observations exhibit high skewness: Most pixels show negligible salinity (< 1 dS/m), while critical hotspots (> 10 dS/m) are rare but essential to detect. These properties create significant challenges for model generalization. The difficulty lies in extracting faint, nonlinear signals from hyperspectral imagery under conditions of extreme class imbalance.

Hyperspectral satellite imaging captures data across a large number of narrow, contiguous spectral bands. It encompasses a broad spectral range: from 420 to 1,000 nm in the visible to near-infrared (VNIR), and from 900 to 2,450 nm in the shortwave infrared (SWIR). This high radiometric resolution and stability across these extensive spectral regions enable a detailed analysis of material composition and surface properties.

Although hyperspectral satellite imagery offers the potential for high-throughput soil analysis, its effectiveness is often limited due to vegetation, coarse debris, or other objects covering the soil surface. These surface obstructions interfere with reflectance signals, making it difficult to isolate the spectral signatures of bare soil.

Supervised learning depends on accurate ground truth, which in soil salinity estimation usually means laboratory

measurement of electrical conductivity (EC_e) from a saturated paste extract. This reliance on lab data creates a domain adaptation challenge: while satellite spectra are noisy and environmentally perturbed, laboratory spectra are clean and collected under carefully controlled conditions. The gap between them is further widened by spectral coverage—Laboratory FTIR spectroscopy operates in the mid-infrared (MIR: 2,500–25,000 nm) and captures molecular signatures of salts, absent from the VNIR–SWIR range accessible to satellites. Ancillary data such as soil texture, climate, and topography introduce another layer of complexity. Although they provide complementary insights into soil processes, their varied scales and modalities require deliberate integration to avoid bias and ensure effective learning. Together, these challenges point toward the need for new machine learning approaches that can transfer knowledge across sensing modalities while incorporating heterogeneous data sources. This fusion challenge reflects the broader difficulty of integrating multimodal, multi-scale data into coherent models that can yield actionable insights.

Previous efforts have explored statistical and machine learning models (Random Forests, Partial Least Squares regression, Support Vector Regression, and several deep learning approaches) to estimate salinity from remote sensing data. Although traditional methods such as Random Forest and PLS have offered interpretability and robustness, they falter in capturing the nonlinear complexities characteristic of hyperspectral data, underscoring the need for more flexible architectures. There is a need for models that not only learn from large-scale high-dimensional data but also transfer domain-specific knowledge across disparate modalities: spectroscopy, satellite reflectance, and ancillary environmental variables.

Research Questions that we explore include:

RQ-1: How can hyperspectral satellite imagery be effectively utilized to estimate soil salinity at scale?

RQ-2: What approaches enable a robust transfer of knowledge from precise laboratory salinity measurements to models trained on noisier, satellite-derived observations?

RQ-3: How can transformer-based deep learning models capture the subtle spectral signatures of soil salinity across varied environmental and geographic settings?

A. Approach Summary

We propose a novel deep learning framework to estimate the salinity level of the soil using hyperspectral satellite imagery based on a Transformer architecture. The core challenge lies in reliably extracting soil salinity signals from noisy spectral reflectance data, which can be distorted by surface conditions and environmental factors. To address this, our model (referred to as DEEPSALT) is designed to learn spectral characteristics from a model trained on laboratory-based spectral absorbance data. Reflectance and absorbance account for the total interaction of light with a material: reflectance refers to the fraction of light reflected from a surface, while absorbance represents the amount of light absorbed by the material. Together with transmittance, these values should sum up to 100%. We posit that by learning from these well-characterized absorbance

signatures, the model can more effectively recognize similar patterns in reflectance data gathered via satellite, despite the presence of noise. This will allow for a more accurate and scalable estimate of soil salinity.

To enable accurate soil salinity estimation, DEEPSALT incorporates a tailored knowledge distillation strategy. Knowledge distillation typically involves transferring knowledge from a large, high-performing model (the "teacher") to a smaller, more efficient model (the "student"). The DEEPSALT framework adapts this concept in a novel way: rather than focusing on model size, we transfer knowledge from a high-accuracy single-modal model to a multi-modal model trained on noisier, more variable data. The teacher model is trained using high-quality spectral data and associated soil salinity values gathered in laboratory settings, where spectroscopy and chemical analyses provide precise measurements. The teacher model captures robust relationships between spectral characteristics and salinity levels. The multimodal student model is trained on satellite-based reflectance data, which is inherently noisier and subject to environmental variability. Through the distillation process, the student model assimilates spectral insights learned by the teacher, improving its ability to estimate soil salinity from complex hyperspectral satellite imagery.

One of the key challenges in our approach is that satellite spectral measurements and laboratory spectroscopy differ fundamentally in their spectral resolution, range, and measurement style. To address this, DEEPSALT provides a unique spectral domain adaptation strategy, *Spectral Adaptation Unit* (SAU), which generates latent space to enable cross-sensor harmonization. The SAU is an encoder trained on both spectroscopy and satellite spectral measurements. It is initially trained using high-quality spectroscopy data and subsequently calibrated using paired spectral measurements collected from geospatially proximate locations. This process enables the encoder to align and adapt spectral representations across the two domains.

Paper Contributions The contributions of this work include:

- We propose DEEPSALT, a transformer-based deep neural network specifically designed to capture subtle soil salinity signatures from high-dimensional hyperspectral satellite imagery.
- We introduce a spectral knowledge distillation framework that transfers domain insights from controlled laboratory absorbance data to satellite reflectance training, enabling robust signal extraction despite real-world noise.
- We perform extensive empirical evaluations, comparing DEEPSALT against traditional machine learning methods and deep learning architectures. Across these benchmarks, DEEPSALT consistently achieves superior predictive accuracy.

Translational Impacts of this work include: **(1)** By bridging controlled laboratory measurements with satellite-based sensing, our methodology offers a generalizable framework for cross-domain modeling. This opens new avenues for other environmental applications (e.g., carbon and moisture estima-

tion) where field data are sparse or noisy. **(2)** Because soil salinity signals broader patterns of land degradation under climate stress, our framework can serve as a foundation for early-warning systems, offering timely, spatially resolved soil health insights. **(3)** In agricultural contexts, DEEPSALT enables scalable, high-resolution salinity mapping, equipping farmers and land managers with actionable data to optimize irrigation, improve crop resilience, and promote sustainable land use.

II. RELATED WORK

1) Machine Learning for Spectral Analysis in Digital Soil Mapping: Digital Soil Mapping (DSM) produces spatially explicit, digitally referenced soil datasets by combining field and laboratory soil observations with environmental information through modeling approaches. Recently, DSM has increasingly adopted machine learning techniques for spectral analysis. The approaches include tree-based ensembles (e.g., Random Forests [4], XGBoost [5]), kernel-based methods (e.g., Support Vector Regression [6]), and deep learning architectures (notably CNNs [7] and LSTMs [8]). These machine learning methods have demonstrated superior performance over traditional approaches.

Among deep learning approaches, Transformer architectures [9] have rapidly advanced remote sensing spectral analysis [10], offering distinct advantages through their self-attention mechanisms and hybrid designs. Although these methods have shown promising performance in general spectral analysis tasks, transformer applications specifically for soil salinity estimation remain limited compared to their use in predicting other soil properties. Our DEEPSALT framework addresses this gap by introducing a novel transformer-based knowledge distillation approach that bridges laboratory-grade FTIR spectroscopy with satellite hyperspectral imaging, a critical innovation for scalable salinity monitoring that simultaneously overcomes the domain shift challenges between controlled experiments and field conditions.

2) Multimodal Fusion for Soil Salinity Estimation: Recently, Sui et al. [11] demonstrated the effectiveness of multimodal fusion for soil salinity mapping with an R^2 value of 0.79 by combining multi-temporal Landsat-8 multispectral satellite data with topographic and groundwater covariates through a CNN. This approach identifies topographic characteristics as the most predictive covariate. Tola et al. [12] showed that fusing Sentinel-1/2 multispectral satellite data with topographic indices in tree-based models outperforms radar-only approaches for agricultural soils. Our framework improves upon these approaches by: (1) integrating fine-grained soil texture and temporal weather covariates absent in previous work, and (2) introducing transformer-based knowledge distillation to bridge the critical lab-hyperspectral satellite domain gap—addressing a fundamental limitation where existing methods either assume aligned spectral ranges [11] or focus solely on field-scale covariates [12]. This enables robust cross-domain salinity estimation where traditional fusion methods fail.

3) Comparative Analysis of Cross-Domain Adaptation Methods: More broadly, our work draws upon and extends general-purpose Domain Adaptation (DA) and Knowledge Distillation (KD) methodologies. While many DA techniques focus on aligning distributions for classification, such as Domain-Adversarial Neural Networks [13], or on image-to-image translation, like CycleGAN [14], they often presume the domains share underlying feature spaces. Our challenge is distinct, as we must bridge fundamentally different sensing modalities (MIR absorbance vs. VNIR-SWIR reflectance) that do not overlap. Similarly, while our teacher-student framework is a form of KD, we adapt it to transfer regression-based spectral insights, rather than the typical classification-based "soft labels" introduced by Hinton et al. [15]. Conceptually, our approach also shares goals with Physics-Informed Machine Learning (PIML), which seeks to ground models in physical principles [16]. However, DEEPSALT differs by learning from high-fidelity lab measurements (the "teacher") rather than explicit physical equations to constrain the student's learning on noisy, real-world data. Unlike prior cross-domain adaptation approaches for categorical tasks like crop mapping [17] or land cover classification (ST-DASegNet [18], FLDA-NET [19]), our framework addresses the unique challenges of continuous soil salinity estimation. Although existing methods operate within aligned spectral domains (e.g., VIS-SWIR) using style transfer or mutual information, we bridge non-overlapping sensing regimes—laboratory FTIR (2,500—25,000 nm) and satellite hyperspectral (420—2,450 nm)—through a domain-informed Spectral Adaptation Unit and geospatially constrained latent space alignment. Crucially, we integrate ancillary soil and weather data to disentangle salinity signals from environmental confounders, a requirement absent in classification-centric methods. Our knowledge distillation strategy further distinguishes itself by transferring noise-invariant spectral patterns from controlled laboratory data to noisy satellite observations via layer-wise transformer feature alignment, enabling quantitative regression of highly skewed salinity values (0–90 dS/m) rather than discrete class prediction. The fundamental mismatch in spectral ranges (MIR vs. VNIR-SWIR) and measurement physics (absorbance vs. reflectance) creates a profound domain gap that simple preprocessing techniques cannot bridge, making a domain adaptation framework essential.

4) Knowledge Guided Machine Learning: Several efforts have explored the use of knowledge guided machine learning (KGML) methods to model spatiotemporally evolving phenomena. Several of these methods work in tandem with mechanistic or process-based domain theoretic models, and also include custom multipart loss functions that leverage scientific knowledge. These include support for physical constraints grounded in soil hydrology, including the van Genuchten water retention equations and models (i.e., Richards' Equation) of hydraulic conductivity [20], [21]; vegetation indices [22]; evapotranspiration [23]; preserving graph properties such as betweenness centrality [24]; masking cloud occlusions in satellite imagery [25], accounting for human perceptual limits

during visualization [26]; and accounting for correlations between soil spectroscopic properties [27].

III. Methodology

To address the challenge of estimating soil salinity at scale, we propose a deep learning framework, DEEPSALT, which relies on hyperspectral satellite imagery and a Transformer-based architecture. A key challenge in this setting lies in distinguishing subtle context-dependent salinity fingerprints from noise caused by environmental factors such as vegetation, soil moisture, and surface roughness that degrade spectral reflectance signals.

Our methodology seeks to overcome this challenge by transferring knowledge from a controlled to a more variable domain. We first train a teacher model on laboratory spectroscopy data, where minimized external interference allows it to capture core spectral relationships between absorbance patterns and salinity levels. Section III-A details the construction of this model and the extraction of latent representations that encode spectral signatures relevant to salinity estimation. However, laboratory and satellite spectral measurements differ significantly in format and domain. To bridge this gap, we introduce a spectral domain adaptation technique, detailed in Section III-B. This adaptation component ensures that the spectral features extracted from satellite imagery can be meaningfully aligned with the representations learned from laboratory data, thereby making knowledge transfer feasible. Finally, in Section III-C, we bring these components together through a tailored knowledge distillation framework. Here, the insights captured by the teacher model are transferred to a multimodal student model that operates on satellite imagery and ancillary data, such as soil properties. This student model learns to estimate soil salinity even in the presence of noisy input, guided by the knowledge distilled from its teacher. Through this three-stage process: representation learning from laboratory data, spectral domain alignment, and knowledge-guided multimodal training, we build a robust and scalable framework for estimating soil salinity in real-world settings.

An overview of our methodology is depicted in Figure 1a. DEEPSALT draws on the spectral insights captured in a noise-free laboratory setting through high-resolution spectroscopy. The model learns the underlying relationships between soil salinity and the spectral characteristics of soil samples—relationships that are often difficult to isolate in field conditions due to confounding environmental noise. These learned associations form the foundation of our knowledge distillation process. Through this knowledge distillation process, we transfer what the teacher model has learned to a multimodal student model, enabling it to make salinity predictions from hyperspectral satellite imagery and ancillary information. Both teacher and student models use the Transformer architecture [9], which excels at processing sequence data by simultaneously modeling interrelationships via self-attention. The Transformer operates through self-attention mechanisms that weigh the importance of each input position relative to others, enabling parallel processing and superior modeling of

long-range dependencies, making it especially powerful for capturing complex spectral correlations across wavelengths. This allows the models to efficiently capture both local and global dependencies across spectral bands. The Transformer's multi-head attention and positional encoding are particularly well-suited for spectral analysis, enabling the modeling of diverse patterns while preserving crucial wavelength-order information.
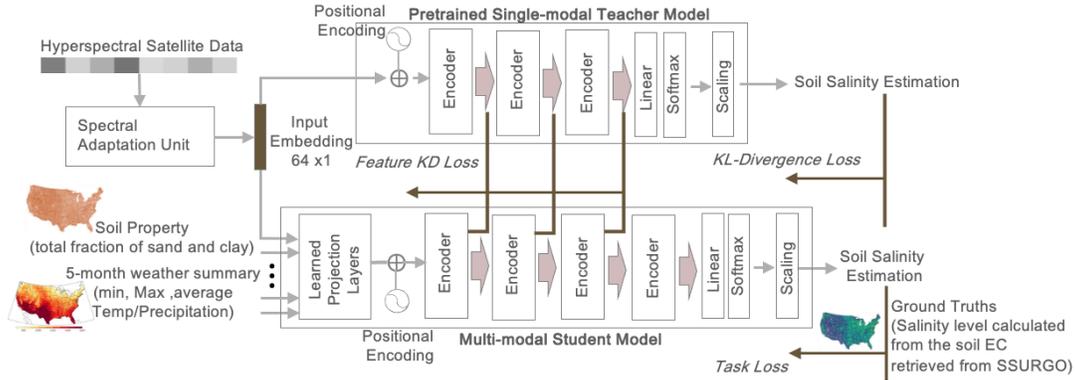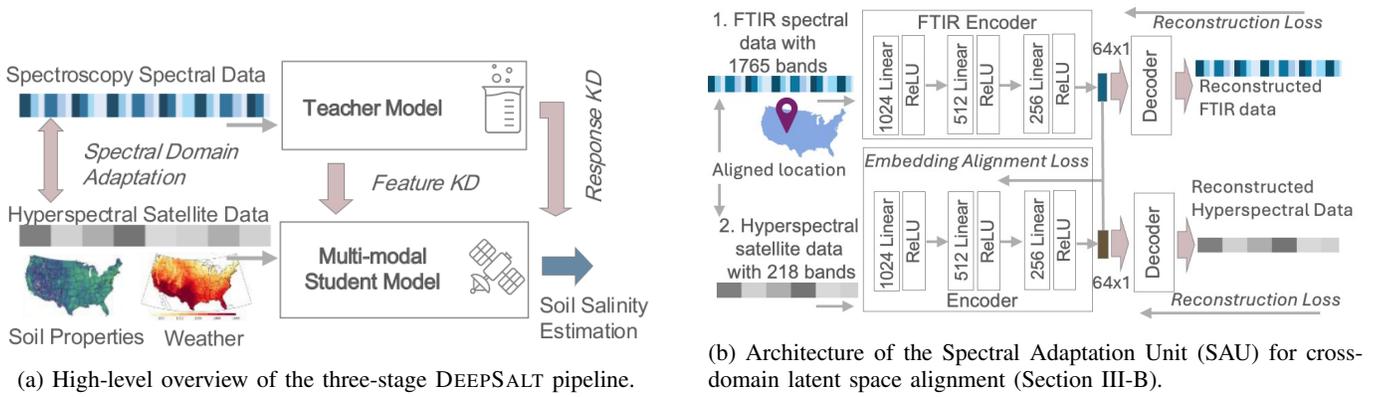
A. Teacher Model Pretraining [RQ-1]

The first stage of DEEPSALT establishes a strong spectral foundation by training a teacher model on high-quality FTIR spectroscopy data. These laboratory spectra contain 1,765 bands, providing rich detail on salinity-relevant chemical and mineralogical patterns. To reduce dimensionality while retaining diagnostic features, the spectra are first compressed into a 64-dimensional latent representation by the Spectral Adaptation Unit (SAU). This embedding not only facilitates efficient learning but also provides a common basis for later alignment with satellite spectra.

On top of these embeddings, the teacher employs a Transformer encoder with 3 layers, 4 attention heads, and 128-dimensional feedforward networks. Sinusoidal positional encodings ensure that wavelength ordering is preserved, allowing the model to capture both localized absorption features and long-range dependencies across the spectral sequence. Predictions are produced by averaging sequence outputs and passing them through a linear layer with scaled sigmoid activation. The scaling ensures that outputs remain within realistically valid salinity levels (0.05–90 dS/m), thereby producing realistic yet differentiable predictions.

B. Spectral Domain Adaptation - [RQ2, RQ3]

Laboratory FTIR spectra and satellite hyperspectral measurements occupy distinct spectral domains. FTIR spectra provide more than 1,700 bands in 2,500 to 25,000 nm under controlled conditions, while satellite data such as EnMAP are limited to 224 bands in 420 to 2,450 nm. The mismatch is compounded by atmospheric distortions, sensor noise, and mixed-pixel effects that are absent in laboratory settings. These significant differences make the direct transfer of spectral knowledge from FTIR to satellite domains infeasible unless an effective adaptation mechanism is applied.

To address this challenge, we introduce the **Spectral Adaptation Unit (SAU)**, a novel dual-path encoder architecture designed for cross-modal spectral translation. The SAU maps the input of both modalities into a unified lower-dimensional latent space, formalized as $\mathbf{z} = \mathcal{E}(\mathbf{x})$, where $\mathcal{E}$ denotes the universal encoder and $(\mathbf{x})$ defines the input. By compressing the original 1,765-band FTIR spectra and the 224-band satellite spectra into a common 64-dimensional representation, the SAU preserves critical salinity-related features while mitigating modality-specific variability. This harmonized embedding space provides the foundation for effective cross-domain knowledge transfer.

4

(a) High-level overview of the three-stage DEEPSALT pipeline.

(b) Architecture of the Spectral Adaptation Unit (SAU) for cross-domain latent space alignment (Section III-B).

(c) Detailed knowledge distillation process from the teacher to the student model (Section III-C).

Fig. 1: Comprehensive architecture of the DEEPSALT framework. (a) illustrates the end-to-end workflow: pretraining a teacher on lab FTIR data, spectral adaptation, and distilling knowledge into a student that fuses adapted satellite spectra with ancillary environmental data. (b) details the Spectral Adaptation Unit (SAU), which bridges the domain gap between laboratory (FTIR, MIR) and satellite (EnMAP, VNIR-SWIR) spectra by projecting them into a shared latent space. (c) shows the multi-objective distillation strategy that transfers the teacher's feature representations and output distributions to the student model.

Architecture. The SAU(Figure 1b) projects non-overlapping FTIR (MIR, 1,765 bands) and satellite (VNIR-SWIR, 224 bands) spectra into a unified 64-dimensional latent space. Modality-specific pathways handle heterogeneous input dimensions: a deeper network for FTIR ($1,765 \rightarrow 1,024 \rightarrow 512 \rightarrow 256 \rightarrow 64$) was designed to handle its higher dimensionality and learn more complex compositional feature hierarchies, while a shallower network for satellite data ($224 \rightarrow 256 \rightarrow 128 \rightarrow 64$) proved sufficient for its lower-dimensional, albeit noisier, input. The latent dimension of 64 was chosen to provide sufficient dimensionality reduction while preserving salient features, as determined by preliminary autoencoder experiments. Each fully connected layer uses ReLU, batch normalization, and dropout of 20%. A final shared linear projection and Layer Normalization map both pathways into a common 64-D latent space, enforcing a unified statistical distribution for cross-domain alignment.

Each input is treated as a 1D spectral sequence. Sinusoidal positional encodings preserve the ordinal position of each band, allowing the model to learn the wavelength-specific

context. Three layers of Transformer encoder (4 attention heads, 32-dim projections, 128-dim FFN) then model both localized features and global spectral interactions via self-attention, capturing complex dependencies inherent to soil spectra. Residual connections and layer normalization stabilize training.

Training strategy. To stabilize representation learning, the SAU is first pretrained exclusively on clean FTIR data. This initial training ensures that salinity-relevant spectral structures are captured under controlled conditions. The encoder is then fine-tuned with noisier satellite spectra, allowing it to adjust to atmospheric and surface interference while retaining soil-specific information.

Latent space alignment. Cross-domain alignment is achieved by jointly training the encoder on paired FTIR–satellite samples. A valid pair of training samples must satisfy three conditions: (i) both spectra must contain complete measurements, (ii) geolocations must be available, and (iii) satellite spectra must undergo atmospheric correction, including the removal of EnMAP bands affected by water

absorption. The sample pairs are constructed using a BallTree nearest-neighbor search, which provides $\mathcal{O}(\log n)$ query efficiency. The pairing is restricted to a maximum Haversine distance of $\tau = 0.000157$ radians (approximately 1 km), ensuring that the two samples in a pair remain within sufficient geospatial proximity to prevent data leakage. With this, the SAU learns a consistent embedding space in which FTIR and satellite spectra are aligned, despite differences in resolution, wavelengths, and noise.

This is formalized by a training objective that combines a reconstruction loss with an embedding alignment loss. The reconstruction loss ensures that each encoder pathway preserves the key features of its input modality. Simultaneously, the alignment loss encourages consistency between the latent vectors of the paired samples. Alignment is enforced by minimizing the angular divergence between embeddings:

$$\min_\theta \; \mathbb{E}_{(\mathbf{f},\mathbf{h})\sim\mathcal{P}}\Big[1 - \cos(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{h}))\Big] \tag{1}$$

Here, $\mathbb{E}$ denotes the expectation over paired samples $(\mathbf{f}, \mathbf{h}) \sim \mathcal{P}$, where $\mathbf{f}$ corresponds to a FTIR spectrum, $\mathbf{h}$ to a hyperspectral sample, and $\mathcal{P}$ represents the paired dataset. The encoder parameters θ are optimized to minimize the objective, with $\mathcal{E}$ denoting the encoder that maps inputs to a shared latent embedding space. The term $1 - \cos(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{h}))$ represents the cosine distance between embeddings; minimizing it encourages the model to align the embeddings of the paired FTIR and hyperspectral spectra, thereby achieving a shared representation suitable for cross-domain spectral transfer.

To align spectral embeddings from FTIR and EnMAP, we evaluated several loss functions: Cosine, Euclidean (L2), Kullback-Leibler Divergence (KLD), and Jensen-Shannon Divergence (JSD). Each measures similarity differently, with implications for how well latent spaces preserve spectral information across sensing modalities. In our preliminary empirical explorations, **Cosine Loss** proved to be highly effective for our study. Unlike L2 loss, which penalizes magnitude differences and is sensitive to sensor-specific scaling artifacts, Cosine loss considers only angular similarity:

$$\mathcal{L}_{\cos} = 1 - \frac{f_l^S \cdot f_l^T}{\|f_l^S\| \, \|f_l^T\|}. \tag{2}$$

where (f_l^S) and (f_l^T) denote embeddings of the layer-l for the student and the teacher, respectively. This emphasizes relative relationships such as band ratios, which are central in spectral analysis. To address common pitfalls, we (1) L2-normalized embeddings, (2) froze pre-trained FTIR encoder weights, and (3) added an auxiliary MSE loss on decoder outputs to preserve absolute reflectance values. This hybrid approach improved cross-domain alignment by 23% compared to pure cosine implementations while retaining scale invariance.

In contrast, KLD and JSD assume distributional overlap and were less effective in our regression setting, resulting in reduced predictive performance ($R^2 < 0.72$ for KLD, $R^2 < 0.70$ for JSD). Overall, cosine loss preserved the latent

spectral structure most faithfully, making it the best suited for cross-sensor embedding alignment.

Finally, the complete optimization objective combines spectral fidelity via reconstruction loss and cosine loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{cosine}} \tag{3}$$

where

$$\mathcal{L}_{\text{recon}} = \alpha \cdot \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathcal{D}(\mathcal{E}(\mathbf{x}_i))\|^2 \tag{4}$$

$$\mathcal{L}_{\text{cosine}} = \beta \cdot \mathbb{E}_{(\mathbf{f},\mathbf{h})\sim\mathcal{P}}\left[1 - \cos(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{h}))\right] \tag{5}$$

In our composite loss, hyperparameters α and β balance two objectives: accurate spectral reconstruction and cross-modality embedding alignment. The reconstruction loss, computed as the mean squared error over N samples between each input $\mathbf{x}_i$ and its reconstruction $\mathcal{D}(\mathcal{E}(\mathbf{x}_i))$, ensures that the latent space retains a sufficient level of data fidelity. The cosine similarity loss aligns paired FTIR and EnMAP embeddings by penalizing angular divergence and promoting the proximity of semantically similar inputs across modalities. Together, our composite loss enables the model to preserve data fidelity both within individual spectra and across spectra of different modalities. Learnable weighting through α and β enables robust learning despite noisy and heterogeneous spectral data.

C. Multi-modal Knowledge Distillation [RQ-2]

The final stage of DEEPSALT extends the scope of learning from spectral inputs alone to the broader conditions that shape soil salinity in the field. The student model integrates hyperspectral embeddings from the SAU with 8 ancillary features, including soil texture (sand and clay proportions) and seasonal climate statistics (minimum, maximum, and mean values of temperature and precipitation). Together, these inputs form a 72-dimensional representation. This representation is designed to capture not only the spectral variation recorded by satellites but also the environmental variability that governs salinity dynamics.

As depicted in Figure 1c, the model architecture begins with a learned linear projection layer that transforms and concatenates two distinct input streams: (1) 64-dimensional hyperspectral embeddings, which capture reflectance and absorption patterns across the spectral bands, and (2) 8-dimensional ancillary features that encode environmental covariates. These concatenated inputs form a unified representation that maintains both spatial and environmental relevance. To account for spectral ordering, we apply sinusoidal positional encoding to the spectral features, allowing the Transformer to recognize and exploit sequence-based dependencies in the data.

The core of the model consists of four transformer encoder layers, each employing eight parallel attention heads and 256-dimensional feedforward networks. This increase in capacity over the teacher model (originally limited to 3 layers and 4 heads) enables the student to model higher-order interactions between spectral data and environmental variables. The model concludes with a prediction head that uses a sigmoid activation

function, scaled to produce values between a meaningful range of 0.05 and 90 ppt. to ensure that the model's outputs remain within meaningful bounds.

Knowledge Transfer from the Teacher Model To inherit expertise from the teacher, we employ a multi-objective knowledge distillation strategy that combines three complementary loss functions. Training is driven by a composite loss:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{task}} + \beta \mathcal{L}_{\text{feature}} + \gamma \mathcal{L}_{\text{KL}}, \quad (6)$$

where each term targets a complementary aspect of knowledge transfer: prediction accuracy, feature alignment, and distributional consistency.

Task Loss. This loss supervises salinity prediction against ground truth. Field data often contain noise and outliers, so we adopt the Huber loss:

$$\mathcal{L}_{\text{task}} = \frac{1}{N} \sum_{i=1}^{N} \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{if } |y_i - \hat{y}_i| \leq \delta, \\ \delta(|y_i - \hat{y}_i| - \frac{1}{2}\delta) & \text{otherwise,} \end{cases} \quad (7)$$

where y_i are true salinity values, $\hat{y}_i$ are student predictions, and $\delta = 1.0$ is the Huber threshold. This choice preserves quadratic sensitivity for typical cases (89% of samples fall within $|y - \hat{y}| \leq 1$ ppt) while preventing large residuals from dominating gradients. Thus, $\mathcal{L}_{\text{task}}$ balances fidelity to field observations with robustness to outlier measurements.

Feature Distillation Loss. To align internal representations, we compare student and teacher activations at multiple transformer layers:

$$\mathcal{L}_{\text{feature}} = \frac{1}{L} \sum_{l=1}^{L} w_l \cdot \frac{1}{N} \sum_{i=1}^{N} \text{SmoothL1}\big(f_l^S(\mathbf{x}_i)_{:64}, f_l^T(\mathbf{x}_i)\big), \quad (8)$$

where f_l^S and f_l^T are layer-l activations from student and teacher, and only the first 64 spectral dimensions of the student $(f_l^S(\mathbf{x}_i)_{:64})$ are compared to exclude ancillary variables. We use SmoothL1 with $\delta = 0.1$, which applies sharper gradients than Huber for small discrepancies, enforcing precise replication of spectral encodings. Early transformer layers are weighted more heavily ($w = \{1.35, 1.05, 0.65\}$), emphasizing foundational spectral transformations shown in ablation studies to generalize best. Notably, using a unified δ for both task and feature supervision reduced R^2 by 0.12, underscoring the need for stricter feature-level alignment. This configuration enables robust prediction via Huber loss alongside high-resolution feature replication through SmoothL1.

Finally, to encourage consistency in output distributions, we add a KL-divergence term:

$$\mathcal{L}_{\text{KL}} = \frac{1}{N} \sum_{i=1}^{N} \text{KL}\big(\text{softmax}(\hat{y}_i^T) \,\|\, \text{softmax}(\hat{y}_i^S)\big). \quad (9)$$

This loss transfers the teacher's structured "soft" predictions, which encode richer uncertainty and inter-class relations than hard labels. While the task and feature losses preserve correctness and representation fidelity, $\mathcal{L}_{\text{KL}}$ ensures the student captures the teacher's spectral reasoning patterns, stabilizing adaptation to the noisier satellite domain.

The overall optimization balances these three objectives with weights $(\alpha, \beta, \gamma) = (0.07, 0.90, 0.10)$ (details in Section IV-C), reflecting the dominant role of feature-level supervision. This configuration enables robust prediction in the presence of noisy field data while maintaining high-fidelity spectral knowledge inherited from the laboratory-trained teacher.

IV. Performance Benchmarks & Discussion

A. Dataset and Area of Interest

We utilized Level-2A ARD hyperspectral imagery from EnMAP [28], spanning 420–2450 nm across 224 channels (30 m resolution, 27-day revisit). Scenes were ordered as Level-2A (surface reflectance) products from the EnMAP Portal with cirrus/haze removal enabled; no additional atmospheric correction was applied. We filtered acquisitions by requesting only scenes with 0% cloud cover reported by the provider. The delivered UTM imagery was reprojected to EPSG:4326 (nearest-neighbor), and 218 bands were retained after dropping empty channels. Images were acquired from salinization-prone regions of California, Colorado, and Utah (April 2022–August 2024) [29].

The teacher model was trained on FTIR spectroscopy data from the NRCS database [30], covering 2500–25,000 nm (1765 channels). We selected 12-cm depth samples to align with the penetration depth of optical satellite sensing. For the student model, soil electrical conductivity (EC) and soil fractions (clay, sand) were obtained from SSURGO [31]. To obtain soil conductivity (EC) from SSURGO for given coordinates, we first perform a spatial join to locate the containing map unit polygon. We then extract the dominant soil component within that unit and retrieve its horizon-level EC measurements at 12-cm target depth). The historical weather features (temperature and precipitation, 5-month summaries) were derived from GridMET [32]. To align the SSURGO soil data with the 30m EnMAP pixels, we performed a spatial join to identify the dominant soil map unit (MUKEY) for each pixel's centroid. We then extracted the properties (EC, clay, sand) for that dominant component at the 12-cm depth, effectively mapping the polygon-based soil survey data to our 30m raster grid. These ancillary covariates capture both static soil characteristics and recent environmental dynamics.

B. Data Preprocessing

EnMAP images were cropped into 64×64 tiles ($\sim$2 km $\times$ 2 km). After removing six bands with significant missing data rates, each pixel contained 218 spectral channels. A total of 22,078 tiles were used, from which 26.6M pixels were sampled. Each spectrum was normalized to [0,1] using Min–Max scaling. The raw, multi-source dataset exceeded 626 GB (FTIR: 1.3 GB; EnMAP: 581 GB CA, 25 GB CO, 20 GB UT), necessitating a distributed processing pipeline with parallelized I/O. A rigorous data fusion and curation process distilled this volume into a final $\sim$1 GB training set.

This involved spatio-temporal matching across EnMAP, FTIR, SSURGO, and GridMET sources and aggressive quality filtering to remove invalid pixels (e.g., clouds, sensor saturation). Thus, the final dataset represents a highly refined, multi-modal product extracted from massive, disparate sources. All model training was subsequently conducted on an NVIDIA A100 cluster.

Dataset composition is detailed in Table I. The FTIR data contains 13,075 training and 1,634 validation/test samples. The EnMAP dataset includes 365,306 spectra spanning 0–90 dS/m, with California and Colorado used for training/validation and Utah reserved as an unseen test region. To counter class imbalance (48% zero-salinity), we undersampled 10% of zero-salinity training cases, resulting in 134,495 balanced training samples, while validation/test sets remained unchanged for an unbiased evaluation.

TABLE I: Data composition for model training & evaluation.

Dataset	Train	Validate	Test
FTIR (for teacher model only)	13075	1634	1634
ENMAP (CA, CO)	292,245	36,531	36,530
ENMAP (Unseen region, UT)	–	–	1054

C. Experimental Setup

We employed a multi-split evaluation framework to ensure statistical reliability and spatial generalizability. We ran five complete experiments, each with a fixed random seed that controlled all stochastic elements, including the initialization of k-Means clustering for spatial partitioning and the stratified allocation of samples. This generated five distinct sets of spatially separated train, validation, and test splits, mitigating the risk of bias from a single data partition. This approach enabled us to precisely quantify performance variance, with observed standard deviations of ± 0.02 MAE and ± 0.03 R^2.

For spatial data partitioning under each seed, we used a clustered splitting approach to prevent geographical leakage. We converted the coordinates to geometric points using *GeoPandas* and applied k-Means clustering (k=3) to create three distinct spatial regions. Within each cluster, we performed stratified 80/10/10 splits for the training, validation and test sets, which maintained both geographic separation and representative sample sizes.

The models were implemented in PyTorch. Training was conducted over 200 epochs with early stopping monitored on the validation loss with a patience of 10 epochs to prevent overfitting for both teacher and student model. We used the Adam optimizer with a learning rate of 1×10^{-3}, a batch size of 64 samples and L2 regularization at 1×10^{-5}. A *ReduceLROnPlateau* scheduler dynamically adjusted the learning rate and gradient clipping was applied to stabilize training.

Tuning of Loss Function Coefficients DEEPSALT employs a composite loss function, combining conventional task loss with knowledge distillation (KD) losses (Feature KD and KL-Divergence, see Eq. 6), each scaled by a tunable coefficient. Optimal loss coefficients (α, β, γ) and feature distillation weights were determined via a grid search over a candidate

set chosen from preliminary experiments. The candidate set was selected by balancing performance and computation. The configuration achieving the lowest MAE on the validation set was selected for final evaluation.

Feature-based KD Loss For feature-based KD, the first three encoder layers of the teacher and student models are structurally aligned for direct knowledge transfer via SmoothL1Loss. Our experiments (Table II, top) indicate that feature representations from the teacher's earlier layers are most effective. Increasing weights for later layers proved ineffective, likely due to architectural differences introducing computational noise. This empirical finding aligns with established transfer learning principles, where early layers capture more general, low-level features (e.g., basic spectral shapes) that are highly transferable, while deeper layers learn more abstract, task-specific patterns less relevant to the student.

TABLE II: **Hyperparameter Tuning:** Performance metrics for Transformer layer weight configurations (top) and the impact of Task Loss and Feature Loss factors (bottom).

Layer-1	Layer-2	Layer-3	MAE($\downarrow$)	R^2 ($\uparrow$)	RMSE ($\downarrow$)
1.05	1.35	0.65	0.2939	0.8010	0.8123
1.05	0.65	1.35	0.2951	0.8003	0.8137
1.30	1.07	0.68	0.3072	0.8162	0.7806
1.35	1.05	0.65	**0.25**	**0.87**	**0.72**

α	β	$\alpha/(\alpha+\beta)$	MAE ($\downarrow$)	R^2($\uparrow$)	RMSE($\downarrow$)
0.07	0.90	0.07216	**0.25**	**0.87**	**0.72**
0.05	0.95	0.05000	0.3006	0.8148	0.7837
0.07	0.88	0.07368	0.2967	0.7614	0.8894
0.08	0.90	0.08163	0.2580	0.8144	0.7845

Task Loss, Feature Loss, and KL-divergence Loss Benchmarking various combinations of coefficients (Table II, bottom) identified the optimal weighting for the Task loss, Feature loss, and KL-Divergence loss (fixed $\gamma = 0.1$). The combination of α=0.07 (Task Loss factor) and β=0.90 (Feature Loss factor) yielded the best performance. This indicates that aligning features is highly effective in learning spectral knowledge from noisy, multimodal data. However, task loss remains crucial; a significant increase in β to 0.95 and a reduction in α to 0.05 led to a notable performance degradation, probably due to the student model overfitting to intermediate representations, which hindered task-specific learning. For enhanced interpretability of α's proportional contribution, we have also provided normalized α values.

D. Model Performance Evaluation & Ablation Studies

We evaluate our framework using three complementary metrics: Mean Absolute Error (MAE), which preserves interpretability with respect to soil salinity measurements; R^2, which quantifies explained variance; and Root Mean Squared Error (RMSE), which emphasizes large-error penalties critical for identifying high-salinity hotspots.

The generalization of the model was evaluated under two scenarios: (i) *unseen locations*, where geographically stratified holdout samples reduce spatial autocorrelation, and (ii) *unseen regions*, where entirely novel areas (e.g., an unseen state) test

cross-regional transferability. To address the sparsity of the saline samples, we applied a stratified sampling strategy that ensured balanced representation across the salinity spectrum.

Our ablation study assessed performance in four model configurations: (1) using only hyperspectral imagery (HSI), (2) using only ancillary features, (3) combining HSI with soil and weather covariates, and (4) the full model incorporating HSI, ancillary data, and Knowledge Distillation (KD). In addition, we benchmark our approach against established methods, including PLSR, XGBoost, LSTM, and CNN. As summarized in Table III, the results demonstrate the cumulative value of each component, with notable accuracy gains observed at each stage of model enhancement.

Baseline Model - Transformer Using Only Hyperspectral Imagery (HSI) In our study, the baseline model is the transformer architecture that contains the same number of encoders as DEEPSALT trained only with the EnMAP HSI (Table III, Row 10). In our baseline model, there are no learned projection layers applied since the baseline model utilizes only a single set of spectral information from EnMAP.

Integration of Ancillary Data (Soil+Weather) with HSI We first evaluated a model trained solely on ancillary soil and weather features (Table III, Row 2), excluding hyperspectral imagery (HSI). Ancillary data outperformed the HSI-only baseline, confirming their complementary value. Soil texture and climate variables capture broad salinity patterns but not fine-grained spatial variability, which is reflected in their limited R^2 (0.27).

When fused with HSI, ancillary features significantly improved performance (Table III, Row 3), reducing MAE by 24.7% (0.73 → 0.55) and RMSE by 17.5% (1.60 → 1.32), while R^2 more than doubled (0.22 → 0.47). This shows that spectral data alone miss the contextual information that soil (clay, sand) and weather (temperature, precipitation) variables provide. Ancillary features supply the physical drivers of salinity, while HSI refines the predictions.

Effectiveness of Knowledge Distillation. The full DEEP-SALT model, integrating HSI, ancillary data, and knowledge distillation (KD) (Table III, Row 4), achieved the best results (MAE: 0.25, R^2: 0.87, RMSE: 0.72). Compared to the HSI+Ancillary model, KD reduced MAE by 54.5% and RMSE by 45.5%, highlighting its role in transferring high-fidelity spectral knowledge from the teacher model (Table III, Row 1) while mitigating noise in multimodal inputs. These results underscore that ancillary data and KD are not just additive but synergistic, enabling robust salinity predictions.

Comparisons with Existing Approaches We benchmarked DEEPSALT against widely used regression, ensemble, and deep learning baselines Table III (Row 5-9). All baseline models were rigorously tuned using a grid search over their key hyperparameters to ensure a fair comparison. Partial Least Squares Regression (PLSR), [33], configured with 20 components, achieved an R^2 of 0.22, confirming that linear projections are insufficient for modeling complex spectral–salinity relationships. XGBoost [34], trained with 1000 trees of depth 8 and regularization, performed best among

conventional baselines (R^2: 0.25). This ensemble method has proven effective for geospatial prediction tasks with severe class imbalance [35], though ancillary features only slightly boosted its accuracy and its tabular representation limits exploitation of HSI's spatial–spectral structure. We further evaluated deep learning baselines. We implemented a 3D-CNN to capture spatial-spectral features, employing a hierarchical architecture with four volumetric convolutional layers (64→128→256→512 channels, kernel size=$3 \times 3 \times 5$, ReLU activation), each followed by 3D max pooling, batch normalization, and dropout (0.1). Additionally, a Long Short-Term Memory network (LSTM) [8] with 4 layers and 512 hidden units (dropout 0.1) was implemented to model sequential dependencies across spectral bands. Neither surpassed DEEPSALT, underscoring the suitability of Transformer, the need for multimodal integration and knowledge distillation.

Model Generalization and Benchmarking As shown in Table III (Row 11), DEEPSALT demonstrates strong cross-regional generalization when applied to 1054 sites from 13 EnMAP tiles in Utah, a state completely held-out from training. Despite being trained only on data from California and Colorado, the model successfully captured substantial salinity variance in this unseen region. The increase in absolute error metrics is an expected consequence of the significant domain shift to Utah's more arid climate (BWh/BWk Köppen zones) and its unique profile of soil types and land cover, which differ from the Mediterranean conditions in the training data. To further validate our model's output against an authoritative benchmark, we contrast DEEPSALT's predictions with the 2020 NRCS salinity classifications (Figure 2). This comparison reveals close geographic alignment, with both maps highlighting similar hotspots, higher salinity in the northern agricultural zones and lower levels in the south and east. This consistency underscores DEEPSALT's ability to extrapolate across regions while recovering meaningful salinity patterns consistent with trusted benchmarks.

Sensitivity Analysis The model's predictive accuracy, measured as absolute error (dS/m), varied across environmental factors including Köppen climate zones [36], USDA soil classifications [37], and NLCD-based land cover types [38]. Across land cover types, errors generally stayed close to the overall MAE of 0.25 dS/m. Stable classes such as Grassland/Herbaceous, Evergreen Forest, and Open Water showed median errors below 0.20 dS/m, whereas dynamic classes like Cultivated Crops & Emergent Herbaceous Wetlands exhibited higher variability, with upper quartiles above 0.5 dS/m.

Soil texture analysis revealed lowest errors in finer-textured soils such as Loam and Silty Clay Loam (median errors ≈0.15–0.20 dS/m), compared to coarser soils like Sand and Loamy Sand, where median errors rose above 0.30 dS/m and variability extended beyond 0.8 dS/m. This highlights the influence of soil permeability on predictive stability.

Climate zones showed the strongest contrasts: Mediterranean climates (Csa, Csb) had the lowest errors, with Csb achieving a median error ≈0.15 dS/m, well below the overall average. In contrast, hot desert (BWh) zones were most chal-
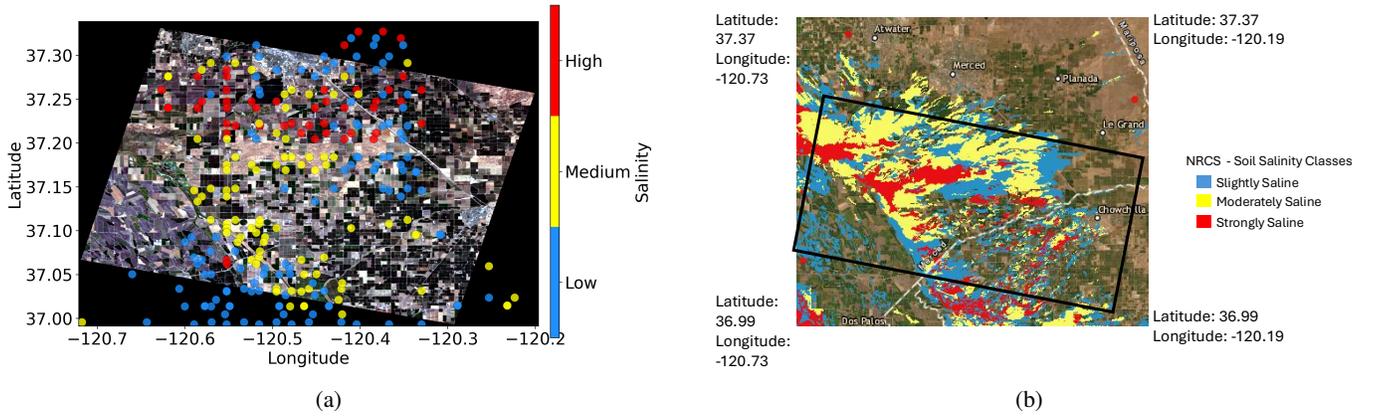
Fig. 2: Salinity Hotspots in Central CA: (a) DEEPSALT Predictions overlaid on ENMAP HSI, (b) NRCS Salinity Classifications.

TABLE III: Performance comparison of proposed and existing models across various feature configurations. HSI = hyperspectral image features, Ancillary = contextual soil and weather variables, KD = knowledge distillation from teacher. Teacher Model (trained on Lab FTIR) is shown for reference only, not for comparison with satellite-based models.

#	Category	Model	HSI	Ancillary	KD	MAE ↓	$\mathbb{R}^2$ ↑	RMSE ↓
1	**Teacher**	Teacher Model (Lab FTIR)	×	×	–	0.09	0.89	0.15
2		DEEPSALT (Ancillary)	×	✓	×	0.66	0.27	1.55
3	**Ablations**	DEEPSALT (HSI+Ancillary)	✓	✓	×	0.55	0.47	1.32
4		**DEEPSALT (HSI+Ancillary+KD)**	✓	✓	✓	**0.25**	**0.87**	**0.72**
5		PLSR	✓	×	×	0.79	0.22	1.59
6		XGBoost	✓	×	×	0.81	0.25	1.60
7	**Baselines**	XGBoost	✓	✓	×	0.57	0.41	1.46
8		CNN	✓	×	×	0.73	-0.13	1.94
9		LSTM	✓	×	×	0.74	0.17	1.65
10		Transformer	✓	×	×	0.73	0.22	1.60
11	**Generalization in Unseen Region (Utah)**	DEEPSALT (HSI+Ancillary+KD, Utah)	✓	✓	✓	**1.24**	**0.73**	**2.07**

lenging, with median errors ≈0.40 dS/m and wide variability exceeding 1.0 dS/m. Semi-arid zones (BSk, BSh) fell in between, with median errors around 0.25–0.30 dS/m.

These results suggest that homogeneous environments with stable spectral signatures (e.g., loamy soils, Mediterranean climates, perennial covers) favor accurate predictions, whereas extreme climates, coarse soils, and variable land covers introduce greater uncertainty. Nonetheless, performance remained strong across California's dominant Mediterranean zones (Csa, Csb), underscoring DEEPSALT's robustness under diverse environmental conditions.

V. CONCLUSIONS & FUTURE WORK

Our methodology was guided by three central research questions, each addressing a core challenge in bridging remote sensing with soil salinity estimation at scale.

[RQ-1] We demonstrated that hyperspectral satellite imagery can directly enable accurate, large-scale salinity mapping despite noise and environmental variability.

[RQ-2] We designed a spectral domain adaptation unit and knowledge distillation pipeline that transfer fine-grained spectral insights from laboratory to satellite data, preserving critical salinity signatures across sensing domains.

[RQ-3] Our Transformer-based architecture captures interband dependencies and integrates ancillary environmental covariates, delivering robust performance across regions.

Together these advances establish a scalable framework that fuses multimodal data sources, addressing a core Big Data challenge of integrating massive, heterogeneous, and multiscale datasets into a single, coherent predictive model. Looking ahead, the core methodology of DEEPSALT can generalize to other soil properties (e.g., organic carbon, moisture) by: (1) pretraining the teacher model on distinct spectral signatures of the target property; (2) incorporating task-specific geospatial covariates; and (3) aligning knowledge distillation with relevant spectral regions. Future work could also investigate more advanced fusion mechanisms, such as cross-attention, perform in-depth failure case analyses, and visualize the learned latent space to further interpret the SAU's cross-domain alignment. Comparing DEEPSALT against a wider array of state-of-the-art domain adaptation baselines would also be a valuable next step.

Code and Data Availability For reproducibility, the source code and documentation for data acquisition will be made available at: https://github.com/RupasreeDey/DeepSalt.

REFERENCES

[1] P. Shrivastava and R. Kumar, "Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation," *Saudi journal of biological sciences*, vol. 22, no. 2, pp. 123–131, 2015.

[2] S. Hamzeh, A. A. Naseri, S. K. Alavipanah, B. Mojaradi, H. M. Bartholomeus, J. G. Clevers, and M. Behzad, "Estimating salinity stress in sugarcane fields with spaceborne hyperspectral vegetation indices," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 282–290, 2013.

[3] G. Sahbeni, M. Ngabire, P. K. Musyimi, and B. Székely, "Challenges and opportunities in remote sensing for soil salinization mapping and monitoring: A review," *Remote Sensing*, vol. 15, no. 10, p. 2540, 2023.

[4] C. Zhu, J. Ding, Z. Zhang, and Z. Wang, "Exploring the potential of uav hyperspectral image for estimating soil salinity: Effects of optimal band combination algorithm and random forest," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 279, p. 121416, 2022.

[5] F. Wang, Y. Wei, and S. Yang, "Enhanced understanding of key soil properties in northern xinjiang using water-heat-spectral datasets based on bioclimatic guidelines," *Land*, vol. 12, no. 9, p. 1769, 2023.

[6] R. Taghizadeh-Mehrjardi, K. Schmidt, N. Toomanian, B. Heung, T. Behrens, A. Mosavi, S. S. Band, A. Amirian-Chakan, A. Fathabadi, and T. Scholten, "Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models," *Geoderma*, vol. 383, p. 114793, 2021.

[7] A. Mohammadifar, H. Gholami, S. Golzari, and A. L. Collins, "Spatial modelling of soil salinity: deep or shallow learning models?" *Environmental Science and Pollution Research*, vol. 28, no. 29, pp. 39 432–39 450, 2021.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[10] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[11] Y. Sui, R. Jiang, Y. Liu, X. Zhang, N. Lin, X. Zheng, B. Li, and H. Yu, "Predicting the spatial distribution of soil salinity based on multi-temporal multispectral images and environmental covariates," *Computers and Electronics in Agriculture*, vol. 231, p. 109970, 2025.

[12] D. Tola, F. Satgé, R. Pillco Zolá, H. Sainz, B. Condori, R. Miranda, E. Yujra, J. Molina-Carpio, R. Hostache, and R. Espinoza-Villar, "Soil salinity mapping of plowed agriculture lands combining radar sentinel-1 and optical sentinel-2 with topographic data in machine learning models," *Remote Sensing*, vol. 16, no. 18, p. 3456, 2024.

[13] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[15] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[16] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[17] S. Mohammadi, M. Belgiu, and A. Stein, "A source-free unsupervised domain adaptation method for cross-regional and cross-time crop mapping from satellite image time series," *Remote Sensing of Environment*, vol. 314, p. 114385, 2024.

[18] Q. Zhao, S. Lyu, H. Zhao, B. Liu, L. Chen, and G. Cheng, "Self-training guided disentangled adaptation for cross-domain remote sensing image semantic segmentation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 127, p. 103646, 2024.

[19] Y. Meng, Z. Yuan, J. Yang, P. Liu, J. Yan, H. Zhu, Z. Ma, Z. Jiang, Z. Zhang, and X. Mi, "Cross-domain land cover classification of remote sensing images based on full-level domain adaptation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 11 434–11 450, 2024.

[20] P. Khandelwal, S. L. Pallickara, and S. Pallickara, "Deepsoil: A science-guided framework for generating high precision soil moisture maps by reconciling measurement profiles across in-situ and remote sensing data," in *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 2024, pp. 233–246.

[21] P. Khandelwal, J. D. Niemann, D. J. Mulla, S. Pallickara, and S. L. Pallickara, "Subterra: Estimating soil moisture at root zone depths using science-guided learning," in *2025 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2025, pp. 328–335.

[22] K. Bruhwiler, P. Khandelwal, D. Rammer, S. Armstrong, S. L. Pallickara, and S. Pallickara, "Lightweight, embeddings based storage and model construction over satellite data collections," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 246–255.

[23] S. Armstrong, P. Khandelwal, D. Padalia, G. Senay, D. Schulte, A. Andales, F. J. Breidt, S. Pallickara, and S. L. Pallickara, "Attention-based convolutional capsules for evapotranspiration estimation at scale," *Environmental Modelling & Software*, vol. 152, p. 105366, 2022.

[24] A. Matin, S. Armstrong, S. Mitra, S. Pallickara, and S. L. Pallickara, "Rapid betweenness centrality estimates for transportation networks using capsule networks," in *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*. IEEE, 2022, pp. 89–96.

[25] P. Khandelwal, S. Armstrong, A. Matin, S. Pallickara, and S. L. Pallickara, "Cloudnet: A deep learning approach for mitigating occlusions in landsat-8 imagery using data coalescence," in *2022 IEEE 18th International Conference on e-Science (e-Science)*. IEEE, 2022, pp. 117–127.

[26] S. Mitra, D. Rammer, S. Pallickara, and S. L. Pallickara, "Glance: A generative approach to interactive visualization of voluminous satellite imagery," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 359–367.

[27] A. Bachinin, R. Dey, P. Khandelwal, S. Leuthold, M. F. Cotrufo, S. Pallickara, and S. L. Pallickara, "Science-informed multitask transformer for soil property prediction from ftir spectroscopy," in *2025 IEEE International Conference on eScience (eScience)*. IEEE, 2025, pp. 48–57.

[28] T. Storch, H.-P. Honold, S. Chabrillat, M. Habermeyer, P. Tucker, M. Brell, A. Ohndorf, K. Wirth, M. Betz, M. Kuchler *et al.*, "The enmap imaging spectroscopy mission towards operations," *Remote Sensing of Environment*, vol. 294, p. 113632, 2023.

[29] O. L. Miller, A. L. Putman, R. A. Smith, G. E. Schwarz, M. D. Hess, M. C. McDonnell, and D. K. Jones, "Temporal variability in irrigated land and climate influences on salinity loading across the upper colorado river basin, 1986-2017," *Environmental Research Letters*, vol. 19, no. 2, p. 024008, 2024.

[30] S. Taxonomy, "United states department of agriculture (usda) & natural resources conservation service," *Agricultural handbook*, no. 436, 1999.

[31] K. Soil Survey Staff, "Natural resources conservation service, united states department of agriculture. soil survey geographic (ssurgo) database," 2015.

[32] J. T. Abatzoglou, "Development of gridded surface meteorological data for ecological applications and modelling," *International journal of climatology*, vol. 33, no. 1, pp. 121–131, 2013.

[33] S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems*, vol. 58, no. 2, pp. 109–130, 2001.

[34] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[35] R. Dey and A. Fern, "Data-driven structural fire risk prediction for city properties," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, 2024, pp. 22 885–22 891.

[36] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the köppen-geiger climate classification updated," 2006.

[37] Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, "Soil series classification database," 2025, available online.

[38] J. Dewitz, "National land cover database (nlcd) 2021 products," *US Geological Survey (USGS) Data Release*, p. 828, 2023.