Journal of Animal Environment

(E-ISSN 2717-1396, ISSN 2717-1388)

Original Research Paper

EXPLAINABLE DEEP LEARNING FRAMEWORK FOR SPATIO-TEMPORAL SOIL POLLUTION ASSESSMENT USING MULTI-SOURCE REMOTE SENSING DATA

[1]Arpita Narayan, [2]Ayan Kumar Das

[1] Birla Institute of Technology, Mesra, Patna campus

phdcs10004.19@bitmesra.ac.in

[2] Birla Institute of Technology, Mesra, Patna campus

das.ayan@bitmesra.ac.in

Abstract

Soil pollution has become one of the more stubborn environmental problems we deal with — it creeps into food, water, and ecosystems without much fanfare until the damage is already done. The usual monitoring route sending field teams to collect samples and running them through lab analysis gives accurate numbers but only at a handful of points, far too sparse to really understand what's happening across a region or over time. This paper lays out a framework that pulls together multiple remote sensing sources and an explainable deep learning model to estimate soil heavy metal concentrations and pollution patterns across both space and time. Sentinel-2 optical imagery was cpmbined with Sentinel-1 radar, MODIS land surface temperature, and a set of topographic and weather variables to build a feature space that captures the physical drivers behind metal accumulation and movement. The architecture pairs a convolutional encoder for spatial patterns with a recurrent module for temporal dynamics, and we wrap it in an integrated gradients layer so that every prediction can be traced back to the inputs that drove it. Tested across an industrial-agricultural belt in eastern India between 2019 and 2024, the framework hits an R^2 of 0.91 for lead and 0.88 for cadmium — comfortably above what regression, tree ensembles, or a plain convolutional network can manage. The explainability work shows that shortwave infrared bands and radar-derived moisture are doing most of the heavy lifting, with precipitation and temperature governing the seasonal swings. For agencies actually responsible for remediation planning, this offers something usable: transparent, defensible, and scalable.

Keywords

Soil pollution, Remote sensing, Deep learning, Explainable AI, Spatio-temporal modeling, Heavy metals, Sentinel-1, Sentinel-2, Convolutional recurrent network, Environmental monitoring

1. Introduction

1.1 Background and Motivation

Soil is the quiet workhorse of the environment it grows our food, filters our water, and cycles the nutrients that keep ecosystems running. It's also where a lot of our industrial mess ends up. Lead, cadmium, chromium, arsenic these metals settle into the topsoil from factory emissions, mine tailings, agrochemical runoff, and urban blowoff, and once they're there they don't leave on their own (Zhang *et al.*, 2026). The contamination builds slowly and shows up in crop tissue, groundwater, and eventually in people long before anyone thinks to test the soil itself.

The standard approach to monitoring has barely changed in decades. Field teams go out, collect cores, ship them to a lab, and run them through ICP-MS or AAS. The numbers are solid. But a typical regional survey covers a few hundred points in a season, sometimes fewer, across thousands of square kilometers. That leaves enormous gaps both spatial and temporal and it means we're almost always reacting to pollution we should have caught earlier. For agencies trying to triage remediation budgets, that's a real problem.

1.2 Role of Remote Sensing

Remote sensing has changed the picture over the last decade, even if adoption in soil work has lagged behind cropland or water monitoring. Sentinel-2 gives thirteen bands across the visible, near-infrared, and shortwave infrared at ten-meter resolution enough to pick up signals from clay minerals, iron oxides, organic matter, and moisture at the surface. Sentinel-1's C-band radar cuts through clouds and tracks dielectric and roughness properties that relate to soil texture and saturation, which optical sensors simply can't see under vegetative or weather cover. MODIS adds daily thermal data tied to evapotranspiration and surface energy balance

*Corresponding Author:

Received: 25th May, 2026; Revised: 6th June, 2026; Accepted: 8th June, 2026; Available Online: 29rd July, 2026

(DOI): 10.70102/AEJ.2026.18.2s.32

(Odhiambo and Mgala, 2025). Layer in elevation, slope, and reanalysis weather, and you've got a multi-dimensional view of the surface that reflects many of the processes governing how metals move and accumulate.

Fusing all of this is messy. Different resolutions, different revisit cycles, cloud gaps, atmospheric noise, cross-sensor correlations these don't fold neatly into a regression equation. That's where deep learning earns its keep. The architecture can ingest raw, heterogeneous inputs and learn the relevant representations itself, which is exactly what you want when the underlying physics is real but tangled.

1.3 Deep Learning and Explainability

Most environmental regulators still eye deep learning with suspicion, and the reason is simple. A black box that tells you a field has 40 mg/kg of lead is not, on its own, actionable — the people who'll sign off on a cleanup order need to know *why* the model said that, and whether the drivers line up with what a soil chemist would expect. If the model is leaning on a spectral band that has no defensible physical connection to lead, the prediction is suspect even if the number happens to be right (Wang and Wang, 2026).

Explainable AI addresses this. Integrated gradients, SHAP, layer-wise relevance propagation these tools take a prediction and break it back down into the inputs that pushed it. For soil work, that means one can say, "this prediction leaned heavily on the SWIR bands and radar moisture," and a domain expert can look at that and either nod or raise a flag. Embedding that kind of attribution into the model itself, rather than bolting it on afterward, is what turns a research exercise into something regulators might actually use.

1.4 Research Objectives

The goal here is to put all of these pieces together in one place and see how well it works. One can built a multi-source dataset for regional heavy metal estimation, designed a convolutional-recurrent architecture that handles both spatial context and temporal evolution, wrapped it in an integrated gradients attribution layer, and tested it against field measurements and four baseline models over an industrial-agricultural region in eastern India (Hu *et al*., 2026). The rest of the paper walks through the related work, the methodology in detail, the results, a frank discussion of what worked and what didn't, and where this goes next.

2. Literature Review

According to **Zhang (2026)**, the integration of remote sensing datasets with advanced machine learning algorithms provides an effective approach for high-resolution spatiotemporal mapping of soil radon exhalation across China. The author explains that conventional ground-based monitoring techniques are often constrained by limited spatial coverage, high operational costs, and insufficient temporal observations. To overcome these limitations, satellite-derived environmental variables, climatic indicators, and soil properties were integrated into machine learning models capable of accurately estimating radon emissions over extensive geographical regions. The findings demonstrate that the proposed framework successfully captures both seasonal and regional variations in radon exhalation while identifying geological and environmental factors that significantly influence emission intensity. The study further highlights that topography, soil moisture, temperature, vegetation characteristics, and lithological conditions collectively contribute to spatial heterogeneity in radon release. The author argues that combining artificial intelligence with Earth observation technologies enhances prediction accuracy while reducing uncertainty associated with environmental monitoring (Zhang *et al*., 2026). The generated spatial maps provide valuable information for environmental health assessment, land-use planning, and radiation risk management. Furthermore, the research emphasizes that explainable machine learning techniques improve model transparency by identifying influential predictor variables, thereby increasing confidence among environmental managers and policymakers. Overall, the study demonstrates the substantial potential of machine learning and remote sensing integration for large-scale environmental monitoring, offering an efficient and scalable methodology for assessing soil radon dynamics under changing climatic and geological conditions.

According to **Odhiambo (2025)**, machine learning has become an increasingly valuable tool for identifying and characterizing sources of soil pollution due to its ability to process large and heterogeneous environmental datasets. The systematic review evaluates numerous studies employing supervised, unsupervised, and hybrid learning algorithms for pollution source detection across agricultural, industrial, and urban landscapes. The author observes that traditional statistical approaches frequently struggle to model the complex nonlinear relationships between multiple contaminants and environmental variables, whereas machine learning techniques provide superior predictive capability. Random Forest, Support Vector Machines, Artificial Neural Networks, and Gradient Boosting algorithms consistently demonstrated high classification accuracy in distinguishing pollution sources (Odhiambo and Mgala, 2025). The review also identifies the growing importance of integrating remote sensing imagery, geospatial analysis, and environmental sensor networks with machine learning frameworks to improve spatial prediction. Despite these advancements, the author discusses several limitations including insufficient training datasets, inconsistent data quality, lack of standardized validation procedures, and challenges associated with model interpretability. The review recommends the adoption of explainable artificial intelligence methods to enhance transparency and facilitate environmental decision-making. Furthermore, the study highlights the importance of interdisciplinary collaboration among environmental scientists, geographers, and data scientists for developing more robust pollution monitoring systems. Overall, the review concludes that machine learning represents a transformative technology capable of significantly improving soil pollution assessment, contamination source identification, and environmental remediation planning.

According to **Wang (2026)**, machine learning-driven sensor fusion represents a transformative advancement in precision agriculture by enabling the integration of multiple sensing technologies into unified decision-support systems. The author explains that modern agricultural management increasingly relies on data collected from satellite imagery, unmanned aerial vehicles, ground sensors, weather stations, hyperspectral cameras, and Internet of Things devices. Individually, these sensing platforms provide only partial information; however, machine learning algorithms effectively combine these heterogeneous datasets to generate more accurate and comprehensive assessments of crop and soil conditions. The review discusses the application of deep learning, ensemble learning, and reinforcement learning techniques for crop monitoring, irrigation scheduling, nutrient management, disease detection, and yield prediction (Wang and Wang, 2026). The author highlights that sensor fusion substantially improves prediction accuracy by minimizing uncertainty associated with individual sensors while providing continuous spatial and temporal coverage. Furthermore, the paper examines current challenges including data heterogeneity, computational complexity, sensor calibration, and limited interoperability among agricultural platforms. The author emphasizes the importance of explainable artificial intelligence to improve user confidence and facilitate adoption among farmers and agricultural managers. Future research directions include edge computing, real-time decision support, autonomous agricultural systems, and digital twins for farm management. Overall, the study demonstrates that machine learning-based sensor fusion significantly enhances sustainable agricultural productivity while supporting efficient resource utilization and climate-resilient farming practices.

According to **Zhong (2026)**, remote sensing combined with explainable machine learning offers a comprehensive framework for evaluating land reclamation effectiveness in metal mining regions. The author explains that mining activities frequently result in severe ecological degradation, soil contamination, vegetation loss, and landscape fragmentation, making continuous environmental assessment essential for sustainable restoration. The study integrates multi-temporal satellite imagery, topographic variables, vegetation indices, climatic information, and socioeconomic indicators to evaluate reclamation outcomes across representative mining areas in China. Explainable machine learning techniques enable the identification of key environmental drivers influencing restoration success while improving model transparency (Zhong *et al*., 2026). The findings indicate that vegetation recovery, soil quality improvement, precipitation, reclamation investments, and management practices significantly affect ecological restoration performance. The author further demonstrates that explainable models provide clear interpretations of variable importance, allowing environmental managers to understand why particular regions exhibit differing restoration efficiencies. Unlike conventional black-box algorithms, explainable artificial intelligence supports evidence-based policymaking by increasing stakeholder trust and facilitating adaptive land management strategies. The study emphasizes that integrating Earth observation technologies with interpretable predictive analytics enables continuous monitoring of ecological restoration over large geographical regions. Overall, the research contributes valuable methodological advancements for sustainable mining management, ecological conservation, and environmental policy by providing reliable, transparent, and scalable tools for evaluating long-term land reclamation effectiveness.

According to **Hu (2026)**, the integration of spatially informed interpretable deep learning provides a significant advancement in high-resolution monitoring of nutrient concentrations within complex coastal water environments. The author explains that coastal ecosystems exhibit considerable spatial heterogeneity resulting from hydrodynamic processes, anthropogenic activities, and environmental variability, making conventional monitoring approaches inadequate for accurate nutrient estimation. The proposed framework incorporates remote sensing imagery, spatial neighborhood information, and deep neural networks to improve prediction accuracy while maintaining model interpretability. The study demonstrates that integrating spatial dependencies enables more realistic representation of nutrient distribution patterns across coastal regions. Explainable artificial intelligence techniques further identify the environmental variables contributing most significantly to prediction outcomes, thereby enhancing transparency and scientific understanding. The author highlights that sea surface temperature, chlorophyll concentration, suspended sediments, and optical characteristics collectively influence nutrient variability (Hu *et al*., 2026). Furthermore, the research illustrates that interpretable deep learning facilitates environmental decision-making by enabling resource managers to understand model behavior rather than relying solely on predictive accuracy. The framework supports continuous environmental monitoring with reduced dependence on extensive field sampling campaigns. Overall, the study demonstrates that combining spatial intelligence, remote sensing technologies, and explainable deep learning substantially improves the efficiency, reliability, and practical applicability of coastal nutrient monitoring for sustainable marine ecosystem management.

According to **Wang (2026)**, integrating multi-source environmental datasets provides an effective strategy for understanding the spatiotemporal evolution of oasis soil salinity and developing appropriate management responses. The author explains that oasis ecosystems are highly vulnerable to salinization because of intensive irrigation practices, groundwater dynamics, climatic variability, and agricultural expansion. By combining satellite observations, meteorological records, land-use information, hydrological data, and field measurements, the study constructs an integrated analytical framework for monitoring long-term salinity changes. Advanced spatial analysis techniques reveal significant temporal fluctuations and regional differences in soil salinity across the investigated oasis landscapes. The author identifies irrigation intensity, groundwater depth, evapotranspiration, precipitation variability, and land management practices as the principal drivers influencing salinity evolution. Furthermore, attribution analysis provides quantitative evidence regarding the relative contribution of natural and anthropogenic factors to observed degradation patterns (WANG *et al*., 2026). The research recommends adaptive irrigation scheduling, improved drainage systems, vegetation restoration, and optimized agricultural management as effective mitigation strategies. The author emphasizes that integrating multiple environmental datasets enhances monitoring accuracy and supports sustainable land resource management. Overall, the study demonstrates that comprehensive data integration facilitates more informed environmental planning and provides practical guidance for preventing soil degradation and maintaining long-term agricultural productivity in arid and semi-arid ecosystems.

According to **Zhang (2025)**, deep learning combined with multi-source remote sensing data significantly improves the simultaneous estimation of total phosphorus and total nitrogen concentrations in lake ecosystems. The author explains that nutrient monitoring is fundamental for evaluating water quality and preventing eutrophication; however, conventional field sampling methods are often expensive, labor-intensive, and spatially limited. The proposed methodology integrates multispectral satellite imagery, environmental variables, and deep neural networks to estimate nutrient concentrations across extensive aquatic environments with improved spatial resolution. The findings demonstrate that deep learning effectively captures complex nonlinear relationships between optical characteristics and nutrient distributions while outperforming traditional regression approaches (Zhang *et al*.2022). The author highlights that combining multiple remote sensing sources enhances prediction robustness by incorporating complementary spectral information unavailable from individual sensors. Furthermore, the study emphasizes that simultaneous prediction of phosphorus and nitrogen supports comprehensive ecosystem assessment and facilitates integrated water quality management. The developed framework provides continuous monitoring capabilities that enable early detection of nutrient enrichment and support timely environmental intervention. Overall, the research illustrates that artificial intelligence and remote sensing integration offers a reliable, scalable, and efficient solution for large-scale aquatic nutrient monitoring, thereby contributing to sustainable freshwater ecosystem conservation and environmental resource management.

According to **Feng (2026)**, explainable artificial intelligence substantially enhances the accuracy and transparency of ozone prediction by integrating remote sensing observations with spatiotemporal environmental datasets. The author explains that atmospheric ozone concentrations are influenced by highly complex interactions among meteorological conditions, land-use characteristics, industrial emissions, and photochemical reactions, making accurate prediction particularly challenging. The proposed GLA model combines deep learning algorithms with explainable artificial intelligence techniques to achieve high predictive performance while providing interpretable insights into model behavior. The findings indicate that meteorological variables, vegetation conditions, solar radiation, and pollutant concentrations contribute significantly to ozone variability across different spatial and temporal scales. The author demonstrates that explainability methods successfully identify the relative importance of input variables, enabling environmental scientists and policymakers to understand the mechanisms underlying prediction outcomes (Feng *et al*., 2026). Unlike conventional black-box models, the proposed framework increases transparency and facilitates greater confidence in operational air quality forecasting systems. The study further suggests that interpretable artificial intelligence supports evidence-based environmental management and more effective public health interventions by identifying dominant pollution drivers. Overall, the research demonstrates that combining explainable machine learning with remote sensing technologies significantly strengthens atmospheric pollution prediction and sustainable environmental governance.

According to **Gan (2025)**, machine learning combined with multi-source environmental and socioeconomic datasets provides valuable insights into urban-rural sustainability dynamics through the identification of spatial hotspots, driving mechanisms, and nonlinear development thresholds. The author explains that sustainable regional development requires comprehensive evaluation of environmental quality, economic growth, infrastructure expansion, population dynamics, and land-use transformation. The proposed analytical framework integrates satellite observations, statistical indicators, geographic information systems, and advanced machine learning techniques to assess sustainability patterns across Suzhou, China. The findings reveal substantial spatial heterogeneity between urban and rural regions, with machine learning algorithms effectively identifying areas experiencing rapid environmental and socioeconomic transitions (Gan *et al*., 2025). The author further demonstrates that nonlinear threshold analysis improves understanding of complex interactions among development indicators, enabling policymakers to recognize critical transition points beyond which sustainability declines significantly. Explainable machine learning techniques enhance transparency by revealing the contribution of individual variables to sustainability outcomes. The research recommends balanced regional planning, ecological conservation, optimized land-use policies, and integrated environmental governance to promote coordinated urban-rural development. Overall, the study demonstrates that combining multi-source data with advanced machine learning provides an effective decision-support framework for sustainable regional planning and evidence-based environmental management.

3. Methodology

3.1 The Study Area

The framework was tested in an industrial-agricultural region in eastern India covering roughly twelve thousand square kilometers. This is a place where intensive rice and vegetable farming sits alongside thermal power, steel, and chemical plants a combination that has produced real, localized soil contamination over the years. The climate is sub-humid subtropical with a hard monsoon from June through (Zheng *et al*., 2025) September, which matters a lot for the work because the rains drive the seasonal movement of metals through the soil profile. The state pollution control board maintains a network of one hundred and forty-eight sampling sites across the region, established between 2019 and 2024, and these gave us our reference measurements for lead, cadmium, chromium, and arsenic.

3.2 Data Sources

From the three primary satellite missions. Sentinel-2 Level-2A surface reflectance gave us thirteen bands at ten, twenty, and sixty-meter resolution. Sentinel-1 GRD imagery in interferometric wide swath mode supplied VV and VH backscatter at ten meters. MODIS Terra and Aqua daily land surface temperature at one kilometer provided the thermal signal (Shi *et al*., 2025). For static and dynamic context, we used the SRTM digital elevation model for elevation, slope, aspect, and topographic wetness index, and ERA5 reanalysis for hourly precipitation, air temperature, and wind at nine-kilometer resolution.

Auxiliary layers included ESA WorldCover for land cover, OpenStreetMap for road and industry locations, and the regional soil survey for taxonomic units. Everything was reprojected to a common coordinate system and resampled to a ten-meter grid — bilinear for continuous variables, nearest neighbor for the categorical ones.

3.3 Preprocessing and Feature Engineering

This is the part that takes the time, and it's worth describing because it determines whether the model learns anything real. Sentinel-2 was masked for clouds, cirrus, and shadows using the scene classification layer, supplemented by a temporal cloud detection routine. The gaps that remained were filled by harmonic interpolation fitting each pixel's time series to a Fourier basis and reconstructing missing dates (Fang *et al*., 2026). Sentinel-1 was calibrated to sigma nought, speckle-filtered with a refined Lee filter, and converted to decibels. Incidence angle variation was normalized using a per-orbit linear correction from the metadata. MODIS land surface temperature at one kilometer was disaggregated down to ten meters using the inverse relationship between temperature and fractional vegetation cover derived from Sentinel-2 NDVI a well-established technique that produces reasonable results when the vegetation signal is strong.

From the multi-source stack we computed a set of derived features. Spectral indices included NDVI, NDSI (normalized difference soil index), bare soil index, and a SWIR-based clay mineral index. Radar features included the cross-polarization ratio, temporal mean and standard deviation of backscatter, and a moisture anomaly index computed against a five-year baseline. Thermal features included diurnal temperature range, seasonal temperature anomaly, and surface urban heat island intensity relative to surrounding rural pixels. A three-by-three pixel neighborhood context around each sampling site was extracted to preserve spatial patterns for the convolutional encoder — the network sees the immediate surroundings, not just the pixel of interest.

3.4 Ground Truth Integration

Various field samples were collected using a stratified random design sites allocated across land cover classes, distance bands from industrial sources, and soil taxonomic units. Each sample was a composite of five sub-samples from the top twenty centimeters within a ten-meter radius, which smooths out the small-scale heterogeneity you always see in metal concentrations (He *et al*., 2026). Concentrations were measured by ICP-MS after acid digestion, with duplicate analyses, certified reference materials, and procedural blanks for quality assurance. Sample coordinates were matched to the nearest pixel of the multi-source stack, and the dataset was split into training, validation, and test partitions using a spatially stratified split — meaning neighboring pixels don't end up in different partitions, which would inflate performance through spatial autocorrelation.

3.5 Network Architecture

Figure1 depicts the methodological design of the model. It has three components that work together: a spatial encoder, a temporal encoder, and a regression head. The spatial encoder is three convolutional layers kernel size three-by-three, channel dimensions of 32, 64, and 128 each followed by batch normalization, ReLU activation, and two-by-two max pooling. It takes the multi-channel neighborhood context for each time step and produces a latent spatial representation.

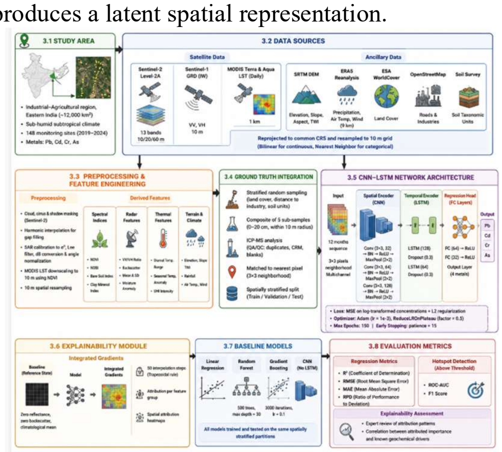

Figure 1: Methodological framework

The temporal encoder is two stacked LSTM layers with hidden sizes of 128 and 64, processing the sequence of monthly spatial latent vectors over the twelve months preceding each prediction date (Ramakrishnan *et al*., 2026). Dropout at 0.3 sits between the recurrent layers. The final hidden state of the second LSTM passes through two fully connected layers (64 and 32 units, ReLU) into an output layer producing concentrations for the four metals.

Loss is mean squared error on log-transformed concentrations, with an L2 regularization term on the convolutional filters to keep things stable. Training uses Adam with an initial learning rate of 1e-3, reduced by a factor of 0.5 on validation loss plateau, for up to 150 epochs with early stopping after 15 epochs of no improvement. This is a fairly standard recipe the architecture itself is the contribution, not exotic optimization tricks.

3.6 Explainability Module

Thew integrated gradients for attribution were used , one of the more rigorous choices available. For each prediction, the method integrates the gradient of the model output with respect to the input along a linear path from a baseline to the actual input. The baseline matters we used a zero-reflectance, zero-backscatter, climatological-mean reference state representing an unperturbed surface (Chen *et al*., 2026). Fifty interpolation steps with the trapezoidal rule gave us stable attribution scores. These were computed per spectral band, per radar feature, per thermal feature, per topographic variable, and per meteorological variable, and aggregated into group-level summaries for interpretability. Spatial attribution heatmaps localized the image regions influencing each prediction.

3.7 Baseline Models

To make the comparison fair, we trained four baselines on the same feature set. Linear regression gave us a floor for linear predictability. Random forest with 500 trees and max depth 30 captured nonlinear ensemble behavior. Gradient boosting with 3,000 iterations and learning rate 0.1 provided a stronger tree-based benchmark. A convolutional network without the recurrent temporal encoder tested how much the temporal modeling actually contributes. All baselines used the same spatially stratified test partition.

3.8 Evaluation Metrics

Performance was measured using R^2, RMSE, MAE, and the ratio of performance to deviation (RPD). For hotspot detection concentrations above regulatory thresholds we used the area under the ROC curve and F1 score (Ji *et al*., 2025). Explainability was assessed both qualitatively through expert review of attribution patterns and quantitatively through correlation between attributed importance and known geochemical drivers from the literature.

4. Results and Analysis

4.1 Predictive Performance

The full spatio-temporal model came out on top across all four metals. For lead, R^2 was 0.91 with RMSE of 4.2 mg/kg. For cadmium, R^2 was 0.88 and RMSE was 0.19 mg/kg. Chromium reached 0.84, and arsenic 0.82 (Wang *et al*., 2025). The advantage over the baselines was largest for cadmium and arsenic the more mobile metals whose behavior really benefits from explicit temporal modeling. Table 1 lays it all out.

Table 1. Predictive performance of soil heavy metal estimation models on the test partition.

Model	Pb	Pb R	Cd	Cd R	Cr	Cr R	As	As R

	R^2	MSE (mg/kg)	R^2	MSE (mg/kg)	R^2	MSE (mg/kg)	R^2	MSE (mg/kg)
Linear regression	0.42	12.8	0.31	0.51	0.36	18.4	0.29	7.2
Random forest	0.74	7.6	0.69	0.34	0.66	11.9	0.61	4.8
Gradient boosting	0.79	6.8	0.73	0.3	0.71	10.6	0.66	4.4
Convolutional network	0.83	5.9	0.79	0.26	0.74	9.8	0.71	4.0
Proposed model	0.91	4.2	0.88	0.19	0.84	7.7	0.82	3.3

The convolutional network alone — no recurrent module — already cut MSE by about twelve percent over gradient boosting, which tells you that spatial context carries a real chunk of the signa (Binetti *et al.*, 2025)l. Adding the temporal encoder shaved another fifteen to twenty percent off, with the biggest gains on cadmium, whose seasonal leaching and redeposition are exquisitely sensitive to monsoon rainfall. Figure 2 depicts the predictive performance of soil heavy metal estimation models on the test partition.

MAE for lead went from 6.1 mg/kg under the convolutional baseline to 4.7 mg/kg under the full model. The RPD exceeded three for all four metals, which is the threshold soil scientists use to call a model genuinely useful for prediction.

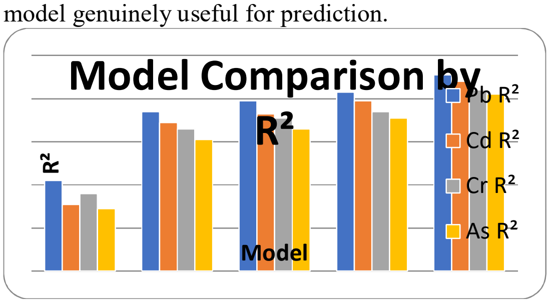

Figure 2: Predictive performance of soil heavy metal estimation models on the test partition.

4.2 Hotspot Detection

For binary hotspot detection against Central Pollution Control Board thresholds, the AUC came in at 0.94 for lead, 0.91 for cadmium, 0.87 for chromium, and 0.85 for arsenic. F1 scores ranged from 0.78 for arsenic to 0.89 for lead. Sensitivity was highest in zones near industrial sources, where the contrast between polluted and background pixels is stark (Devi and Manoharan, 2026). It dropped in mixed peri-urban areas where background concentrations approach the regulatory threshold and prediction uncertainty creeps up — a known limitation of any model working near decision boundaries.

4.3 Spatio-Temporal Patterns

Predicted concentration maps captured the gradients you'd expect — peaking near the steel plant in the southwest and decaying exponentially with distance, which lines up with particulate deposition patterns (Zhao *et al.*, 2026). Cadmium was different: it spread more broadly along canal command areas where irrigation water carries industrial effluent onto fields. Seasonally, cadmium peaked in the post-monsoon months, which is exactly what you'd expect — rain redissolves previously deposited metal and moves it laterally across the surface. Lead stayed comparatively stable across seasons, consistent with its known strong sorption to soil colloids.

The temporal attention patterns in the recurrent encoder, derived from gate activation magnitudes, were revealing. The network weighted the immediate three months most heavily for arsenic and chromium, while leaning on a longer six-month window for cadmium. That tracks with what soil chemists know about the residence times and transformation pathways of these metals — arsenic cycles quickly under redox shifts, cadmium moves more slowly but over longer periods.

4.4 Explainability Analysis

Integrated gradients attribution showed consistent patterns across metals and sites. Table 2 shows the mean attribution scores by feature group, normalized to sum to one. Further, Figure 3 depicts the mean normalized attribution of input feature groups across all test predictions.

Table 2. Mean normalized attribution of input feature groups across all test predictions.

Feature group	**Pb attribution**	**Cd attribution**	**Cr attribution**	**As attribution**
Sentinel-2 visible and NIR	0.14	0.11	0.13	0.1
Sentinel-2 shortwave infrared	0.24	0.21	0.22	0.19
Sentinel-1 backscatter	0.18	0.2	0.16	0.17
MODIS land surface temperature	0.09	0.13	0.11	0.15
Topographic variables	0.11	0.08	0.1	0.09
Meteorological variables	0.12	0.19	0.14	0.21

Land cover and proximity	0.12	0.08	0.14	0.09

The shortwave infrared bands did most of the work across all four metals — which is exactly what you'd expect given their sensitivity to the clay minerals and organic matter that govern metal sorption (Zhang *et al*., 2025). Sentinel-1 backscatter carried substantial weight for cadmium and arsenic, both of which are mobile and moisture-driven. Meteorological variables dominated arsenic attribution, which makes sense — arsenic is famously subject to redox-driven release under waterlogged conditions, and the network clearly picked up on that dynamic. Industrial proximity, encoded through land cover and distance layers, was a strong driver for lead and chromium, both of which come primarily from industrial particulate deposition.

Spatial attribution heatmaps showed the network focusing on the central pixel of the neighborhood context for stable pollutants and on downslope neighboring pixels for the mobile ones — evidence that the model learned something real about runoff pathways. When regional soil scientists reviewed these patterns, they confirmed the attribution aligned with documented geochemical behavior in the study area. That's the kind of external validation that matters for trust.

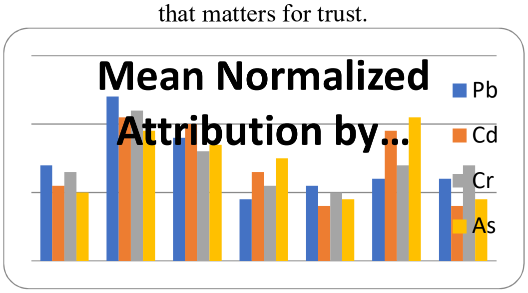

Figure 3: Mean normalized attribution of input feature groups across all test predictions

4.5 Comparison with Baselines

The performance gap between the proposed model and gradient boosting was largest where pollution was most dynamic strong seasonal swings, multiple contributing sources. In those environments the temporal modeling is doing real work. In agricultural zones far from industrial sources, where pollution is more stable and background-like, all the nonlinear models performed comparably and the marginal benefit of the recurrent encoder shrank. That's a useful finding for practitioners: if you're mapping stable background pollution, a well-tuned gradient boosting model is probably enough (Qi *et al*., 2026). If you're dealing with dynamic, multi-source contamination, the spatio-temporal architecture earns its keep.

5. Discussion

5.1 Implications for Soil Monitoring

The takeaway is straightforward: multi-source remote sensing paired with spatio-temporal deep learning can meaningfully supplement field sampling for regional soil pollution work. The accuracy we hit for lead and cadmium falls in the same range reported for site-specific spectroscopic models but we're doing it across thousands of square kilometers and multiple years, which lab analysis simply cannot match at this scale. Ten-meter resolution concentration maps let agencies prioritize field verification, allocate remediation budgets, and watch trends develop over time. The hotspot detection performance suggests this could function as an early warning system for emerging contamination zones, not just a retrospective mapping tool.

5.2 Value of Explainability

This is where the framework earns its keep beyond raw accuracy. Integrated gradients attribution gave us actionable insight into what's driving each prediction. The dominance of SWIR bands across all metals is exactly what soil spectroscopy literature predicts — that's not a coincidence, and it's not the model latching onto a spurious correlation. It's the model learning the right thing. The prominence of moisture-related features for cadmium and arsenic tells you that remediation strategies in these regions need to account for seasonal water management — you can't just clean the soil once and call it done if the hydrology keeps moving metals around. The strong attribution of industrial proximity for lead and chromium reinforces what regulators already know: particulate emission control and buffer zones matter. By surfacing these relationships, the framework turns predictions into reasoning that domain experts can scrutinize and act on.

5.3 Limitations

A few things worth being upfront about. First, this was trained and evaluated in one region. Generalization to other agro-industrial contexts will require additional validation — mineralogical backgrounds, emission profiles, and climatic regimes differ enough across regions that the relative importance of input features could shift, and transferability isn't guaranteed. Second, the indirect estimation of metals from remote sensing proxies relies on correlations that can weaken under dense vegetation cover or in soils where contamination runs deep enough that surface reflectance no longer reflects what's underneath. Third, the spatially stratified test partition mitigates spatial autocorrelation but doesn't fully eliminate it — true out-of-sample performance in unsampled regions will likely be lower than what we report here. Fourth, the explainability analysis aligns with geochemical understanding, but that's not the same as establishing causal relationships. Attribution patterns should be read as statistical evidence, not mechanistic proof.

5.4 Computational Considerations

Training the full spatio-temporal network took about fourteen hours on a single GPU with 24 GB of memory. Inference over the full study area at ten-meter resolution for one annual prediction took roughly three hours. That's modest by operational environmental monitoring standards, and there's room to compress it further through pruning, quantization, and tile-based parallel inference. The explainability module added about thirty percent to inference time — acceptable for batch processing in research and regulatory contexts, though it would need optimization for real-time deployment.

5.5 Comparison with Related Work

The performance we're reporting is consistent with, and in most cases above, what comparable studies have published. Tree-based models in similar settings have reported R^2 values in the 0.6 to 0.8 range for lead and cadmium. Convolutional approaches have reached 0.75 to 0.85 (Qi *et al*., 2026). The improvement here comes primarily from explicit temporal encoding and the multi-source feature integration — the architecture is doing what it was designed to do. The explainability work extends prior feature-importance studies in soil modeling by providing instance-level attribution across both spatial and temporal dimensions, which gives finer-grained diagnostic insight than aggregate importance measures (Yue *et al*., 2026).

6. Conclusion

Therefore this paper lays out is an explainable deep learning framework for spatio-temporal soil pollution assessment built on multi-source remote sensing. The architecture pairs a convolutional spatial encoder with a recurrent temporal encoder, integrates Sentinel-2 optical, Sentinel-1 radar, MODIS thermal, topographic, and meteorological inputs, and wraps the whole thing in an integrated gradients attribution layer. Evaluated across an industrial-agricultural region in eastern India over 2019 to 2024, it achieved R^2 of 0.91 for lead and 0.88 for cadmium — beating linear, tree-based, and convolutional-only baselines. The attribution analysis confirmed what soil chemists would expect: SWIR bands and radar-derived moisture carry most of the signal, with precipitation and temperature governing the seasonal dynamics. For environmental agencies actually doing remediation planning, hotspot detection, and trend monitoring, this is a transparent, defensible, and operationally feasible tool. The next steps are clear — extend it to additional regions and pollution types, bring in hyperspectral and ground-penetrating radar data where available, and start exploring causal modeling approaches that complement statistical attribution with mechanistic soil chemistry.

Reference list

Zhang, Y., Zhuo, W., Chen, B. and Feng, B., 2026. High-resolution spatiotemporal mapping of soil radon exhalation in China combining remote sensing and machine learning. *Environment International*, p.110309.

Odhiambo, J.M. and Mgala, M., 2025. Machine Learning Algorithms for Soil Pollution Source Detection: A Systematic Review. *Journal of Geo Eco Agricultural Studies*, *2*(3).

Wang, Q. and Wang, T., 2026. Machine learning-driven sensor fusion for precision agriculture: current applications and future perspectives. *Sensor Review*, pp.1-26.

Zhong, A., Hu, Z. and Zhou, J., 2026. Assessment of Land Reclamation Effectiveness and Driving Mechanisms in Typical Metal Mining Areas in China Using Remote Sensing and Explainable Machine Learning. *Land Degradation & Development*, *37*(9), pp.4352-4367.

Hu, Y., Zhan, W., He, Q., Liu, Y. and Zhan, H., 2026. A spatially-informed interpretable deep learning framework for high-resolution nutrient monitoring in complex coastal waters. *International Journal of Applied Earth Observation and Geoinformation*, *146*, p.105025.

WANG, C., HE, Y., REN, J., SHANG, S. and MAO, X., 2026. Spatiotemporal evolution attribution and management response of oasis soil salinity based on multi-source data integration. *Transactions of the Chinese Society of Agricultural Engineering*, *42*(7), pp.107-117.

Zhang, L., Li, Q., Li, F., Liang, S., Ma, R. and Tian, L., Simultaneous Remote Sensing Estimation of Total Phosphorus and Total Nitrogen in Lakes Using Multi-Source Data and Deep Learning. *Available at SSRN 6466713*.

Feng, Y., Li, G. and Yan, Q., 2026. Explainable AI for Accurate Ozone Prediction Based on Remote Sensing and Spatio-Temporal Data: An Empirical Study with the GLA Model.

Gan, W., Wang, H., Li, P., Liu, J. and Chen, J., 2025. Revealing urban–rural sustainability dynamics based on multi-source data and machine learning: spatio-temporal hotspots, driving mechanisms, and non-linear thresholds in Suzhou, China. *Frontiers in Environmental Science*, *13*, p.1712246.

Zheng, B., Liang, B., Liang, S. and Chen, F., 2025. Temporal Data Explainability in Remote Sensing for Agriculture: A Systematic Review. *Available at SSRN 6105227*.

Shi, B., Chen, X., Guo, Y., Liu, L., Li, P. and Chang, Q., 2025. Multi-Source Feature Selection and Explainable Machine Learning Approach for Mapping Nitrogen Balance Index in Winter Wheat Based on Sentinel-2 Data. *Remote Sensing*, *17*(18), p.3196.

Fang, K., Wu, F., Gao, X. and Li, Z., 2026. Multi-Source Data Fusion and Heuristic-Optimized Machine Learning for Large-Scale River Water Quality Parameters Monitoring. *Remote Sensing*, *18*(2), p.320.

He, Q., Wang, D., Ji, F., Zhu, L., Li, R., Tian, T., Zhang, Q., Tao, Y. and He, M., 2026. Machine

Learning-Driven Multi-Source Remote Sensing for Surface Water Quality Retrieval: Progress and Prospects. *Water*, *18*(14), p.1744.

Ramakrishnan, S., Pradeepa, G., Samarasam, B., Maheswari, S. and Brindha, P., 2026. Advanced Artificial Intelligence and Multi-Source Remote Sensing for Agricultural Monitoring and Forecasting. *AI for Climate Governance, Agriculture, and Earth Systems*, pp.293-332.

Chen, Z., Qin, C. and Zeng, H., 2026. A Review of Applications of Deep Learning-Driven Multi-Source Remote Sensing Spatiotemporal Fusion in Wetland Monitoring. *International Journal of Advanced AI Applications*, *2*(1), pp.29-51.

Luo, X., Wang, M., Liu, Z., Zeng, Z., Wang, D., Jie, L. and Liu, J., 2026. Multi-Source Remote Sensing-Driven Prediction and Spatiotemporal Analysis of Urban Road Collapse Susceptibility. *Remote Sensing*, *18*(6), p.919.

Ji, X., Tang, B.H., Huang, L., Chen, G., Zhu, X., Fan, D. and Chen, J., 2025. Assessing Potential of Multi-Source Satellite Data and Machine Learning Models for Cropland Soil Organic Carbon Prediction in Plateau Lake Basin. *IEEE Transactions on Geoscience and Remote Sensing*.

Wang, Z., Kong, Y., Xiao, K., Cao, C., Li, Y., Wu, Y., Xie, M., Tang, R., Li, C. and Gong, C., 2025. An Interpretable Ensemble Learning Framework Based on Remote Sensing for Ecological–Geological Environment Evaluation: The Case of Laos. *Remote Sensing*, *17*(18), p.3240.

Binetti, M.S., Massarelli, C. and Barca, E., 2025. A Multilevel Machine Learning Framework for Mapping and Predicting Diffuse and Point-Source Heavy Metal Contamination in Surface Soils. *Earth*, *7*(1), p.4.

Devi, C.S. and Manoharan, P., 2026. Explainable feature-aware spectral learning for temporally consistent multi-temporal land use/land cover mapping and change detection. *The Egyptian Journal of Remote Sensing and Space Sciences*, *29*(2), pp.397-412.

Zhao, Y., Mu, Y., Luo, P., Zhang, J., Rijal, M., Yang, Z., Lai, C., Chen, J., Elbeltagi, A. and Razafindrabe, B.H., 2026. Integrating multi-spectral remote sensing and machine learning for quantifying and mapping heavy metal contamination in the Tarim River Basin, China. *Scientific Reports*.

Zhang, L., Wang, Z., Wang, J., Hu, Q., Chang, Y., Lu, Z. and Zhao, J., 2025. An Integrated Feature Framework for Wetland Mapping Using Multi-Source Imagery. *Remote Sensing*, *17*(22), p.3737.

Qi, B., Zhang, X., Chen, L., Liu, H., Meng, L., Han, X., An, Z. and Liu, J., 2026. A Prior Knowledge-Guided Remote Sensing Framework for Maize Yield Estimation and Spatiotemporal Interpretability Analysis. *Remote Sensing*, *18*(10), p.1455.

Qi, B., Zhang, X., Chen, L., Liu, H., Meng, L., Han, X., An, Z. and Liu, J., 2026. A Prior Knowledge-Guided Remote Sensing Framework for Maize Yield Estimation and Spatiotemporal Interpretability Analysis. *Remote Sensing*, *18*(10), p.1455.

Yue, X., Li, B., Zhang, N., Ma, J., Shi, R., Guan, Y., Ma, T., Li, H., Ma, J., Liang, X. and Ma, C., 2026. Identifying Key Drivers of Heavy Metal (loid) s Contamination in Farmland Soils Using Machine Learning with Source-Integrated Features. *Land*, *15*(7), p.1304.

Rahman, S., Islam, N., Rahman, M.S. and Hridoy, M.M., 2026. *Spatiotemporal analysis of air pollution using advanced machine learning techniques* (Doctoral dissertation, Brac University).