



Analyzing Soil Pollution by Image Processing and Machine Learning at Contaminated Agricultural Field

Dr. Priya Vij ^{1*} , Patil Manisha Prashant ² 

^{1*} Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.

E-mail: ku.priyavij@kalingauniversity.ac.in

² Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.

E-mail: patil.manisha@kalingauniversity.ac.in

Abstract

Due to the fast advancement of big data, applying Machine Learning (ML) techniques to detect Soil Pollution (SP) at Potentially Contaminated Sites (PCS) across many sectors and regional sizes has emerged as a prominent research focus. The challenges in acquiring essential indices of SP sources and routes result in present methodologies exhibiting low predictive accuracy and an inadequate scientific foundation. This study gathered environmental data concerning heavy metal and organic contamination from 200 PCS across six representative sectors. Twenty-one indices derived from fundamental data, potential SP from products and materials, SP efficacy, and the migrating capability of SP were employed to build the SP detection index method. The research integrated the score into the new characteristic group, including 11 indicators using consolidation computation. The newly selected feature subset was utilized for training ML designs, including Random Forests (RF), Support Vector Machines (SVM), and Multilayer Perceptrons (MLP), and evaluated to ascertain its impact on SP recognition methods. The study findings indicated that the four newly developed indices by feature fusion exhibit an association with SP comparable to that of the original index. The component analysis suggests that several indices related to fundamental information, contamination potential from products and raw materials, and SP prevention levels significantly influence SP to varying extents. The index of the migratory capability of soil contaminants has minimal influence on the classification job of SP detection inside PCS. This research introduces a novel technological approach for identifying SP via big data and ML techniques while offering an overview and scientific foundation for PCS's environmental administration and SP mitigation.

Keywords:

Soil pollution, image processing, machine learning, agriculture.

Article history:

Received: 07/06/2024, Revised: 02/08/2024, Accepted: 09/09/2024, Available online: 30/10/2024

Introduction and Background

Due to rapid urbanization and economic development, several industrial businesses have ceased operations or relocated, resulting in many Potentially Contaminated Sites (PCS) (Obiri-Nyarko et al., 2021). Detecting Soil Pollution (SP) is a crucial requirement for the ecological risk analysis of PCS and for implementing ecological oversight of lands utilized by businesses (Askari et al., 2020; Danková et al., 2021). Due to the intricate origins of pollutants, the sample survey methodologies and detection analyses employed to ascertain the SP of PCS frequently yield ambiguous border determinations, significant prediction discrepancies, and elevated expenses. The causes and processes of SP, encompassing source-sink relationships and the extent of the effect of each identifying indicator, must be more adequately delineated. Scientists want to develop an SP recognition system by establishing a pollution recognition index structure, which can be integrated with site data and contamination migration algorithms to more swiftly and effectively identify site pollution and associated risks (Li & Sun, 2024; Paul et al., 2020).

Due to significant variations in collection sources and techniques, mining analytical perspectives, and processing techniques for site environmental information, the index methods employed in prior research to identify SP varied. Previous research primarily utilizes geography, company scale, industry classifications, and manufacturing history to develop index systems (Vinante et al., 2021). In contrast, methodologies for identifying SP from PCS are derived from firm web pages and records between additional resources. Soil classifications, physicochemical characteristics, remote sensing facts, and other fundamental data are critical indicators for identifying SP and genetic pathways (Maurya et al., 2020). Utilizing a review, land usage classifications, and additional data, Baier et al. employed evaluation to assess the impact of metal mines on adjacent SP across several scenarios (Baier et al., 2022). Zhai et al. employed high-resolution mapping imagery, land usage facts, soil classifications, and ecological variables to examine spatiotemporal variations in China's energy extraction and utilization sector (Zhai et al., 2021). Prior research frequently emphasizes the indices and the magnitude of learning information while neglecting contamination processes and driving variables. Detecting SP from PCS at regional dimensions and across many sector categories must have adequate accuracy and scientific rigor (Wood & Blankinship, 2022).

Despite significant advancements in prior research utilizing Machine Learning (ML) techniques for SP identification, the emphasis was mainly on model assessment metrics (Huang et al., 2023). A study on the interpretability of models and the determinants of SP still needs to be completed. Model comprehension is crucial for recognizing SP, as it informs feature engineering, data gathering, and decision-making on the sustainability and prevention of PCS (Elizondo-Martinez et al., 2020). The pollution source-route-acceptor model for companies has four index groups: environmental risk control threshold, SP state, SP migratory route, and SP receptors. This technical definition is a crucial foundation for detecting SP from PCS across various sectors and regional sizes in China (Hu et al., 2021). The technical description employs the expert scoring technique to establish the index rating, which entails a degree of subjectivity. Developing empirical and objective index weighting and scoring standards is essential in choosing feature methodologies and comprehension (Ramesh & Sanampudi, 2022).

Feature selection approaches are employed to investigate the determinants of SP, namely the inherent connection of simple, variable in nature, or tree-oriented designs, to ascertain the significance and featuring ratios of characteristics. For instance, Random Forests (RF) (Van Der Westhuizen et al., 2023), XGBoosts (Ye

et al., 2023), and the remaining modeling choice of feature analysis techniques are employed to determine the significance of SP detection indices. The elucidation of "black box" designs, including Support Vector Machines (SVM) (Deiss et al., 2020) and Convolutional Neural Networks (CNN) (Ray et al., 2020), together with the assessment of the influence of SP detection indices on-site contamination and their beneficial and detrimental correlations, remains insufficient.

Jas et al. introduced the Shapley Additional exPlanations (SHAP) structure, which evaluates every member's involvement in a cooperative function and offers a goal allocation of advantages (Jas & Dodagoudar, 2023). ML defines the issue as a task if characteristics are considered individual participants. By integrating this with ML, the significance of each feature's participation is quantified as the feature's significance.

In contrast to the inherent interpretation approaches of linearised and tree-based method embeddings, the suggested approach is traditional. It generates a score for SVM, Multiple Layered Perceptrons (MLP), and neuronal networks to assess the impact of each parameter (Mosavi et al., 2021). The paradigm offers forecasts that exhibit favorable or adverse correlations with the targeted factor, facilitating regional and universal interpretations (Camgözlü & Kutlu, 2023).

This study developed an SP detection index method encompassing primary data, possible SP from products and raw materials, SP control levels, and the migratory capability of soil contaminants (Khan et al., 2021). In contemporary ecology and environmental research, RF, SVM, and MLP models were employed to develop identification algorithms for SP in PCS. This work aims to (1) validate the precision of ML algorithms developed using pollutant sources and route indices of locations. Identify an attribute fusion technique to enhance the efficacy of ML predictions and the understanding of models utilizing the SHAP architecture. Identify the critical criteria to provide a scientific foundation for directing SP management in factories and enhancing the existing technical standards for SP risk assessment of PCS (Li et al., 2023). This research aims to introduce a novel technological approach for identifying SP via big data and ML techniques while offering a scientific foundation for the ecological management of sites and SP mitigation (Angin et al., 2020; Mohamed et al., 2024).

Materials and Methods

Figure 1 summarizes the three-layer model created for estimating SP risk levels. The initial layer provided comprehensive data on SP. In the secondary level, the picture was disaggregated into elements, and the characteristics of the elements indicating the selected elements were retrieved. The factors are categorized into three categories: 1) The values of red-green-blue (R, G, and B) and their composite index; 2) the lengths and gradients of elements relative to the target terrain, like plants, lakhs, and industries; and 3) The distances and gradients of pixels concerning certain functional sections of manufacturing plants, such as waste from industries disposal zones. The risk for lead SP at the sample stations was selected as a dependent factor. In the subsequent level, many methods, namely RF, Extreme RD (ERF), SVM, and MLP were learned using the obtained features, and their efficacy was assessed. The methodologies pertinent to each component are delineated in the subsequent sub-sections.

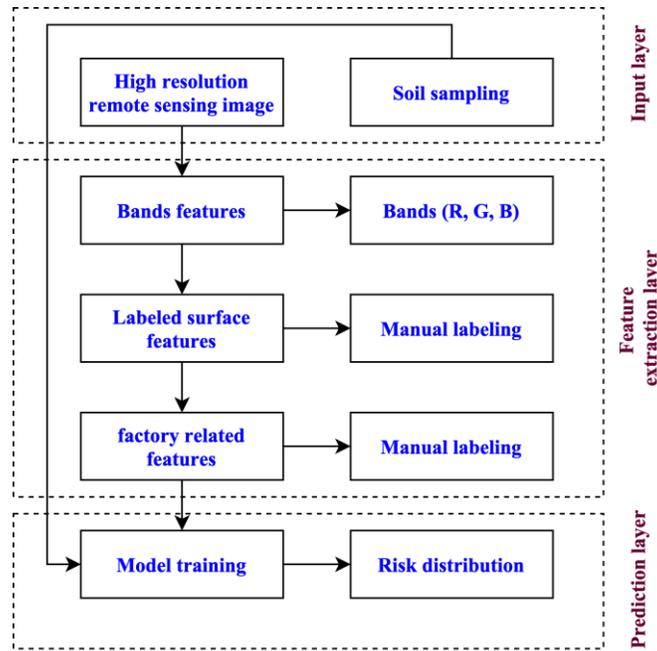


Figure 1. Workflow of the research

Input

The research site is in Zhongxiang (Figure. 2). The environment is monsoon, with an average yearly temperatures of 16.2 °C and a yearly rainfall of 947.2 mm. The average yearly wind velocity is 3.4 m/s, with the predominant wind path being from South-east to North-west. During its formation, this region constituted an ancient sea. After the Silurian period, it was elevated to terrestrial due to the Caledonian orogeny and incorporated into Dahong. During the area, the Himalayan tectonic activity caused variations in elevation and fractures, leading to the development of a framework and the structural structure characterized by an anticline and minor mistakes inside foldings. The most significant thickness ranges from 7160 to 10265 meters. Paddy and flavor-aquic soil in the plain constituted 95.21% of the entire agricultural area. In comparison, the soil layer of the steep hills included approximately 4.22% of the farm area in the city.

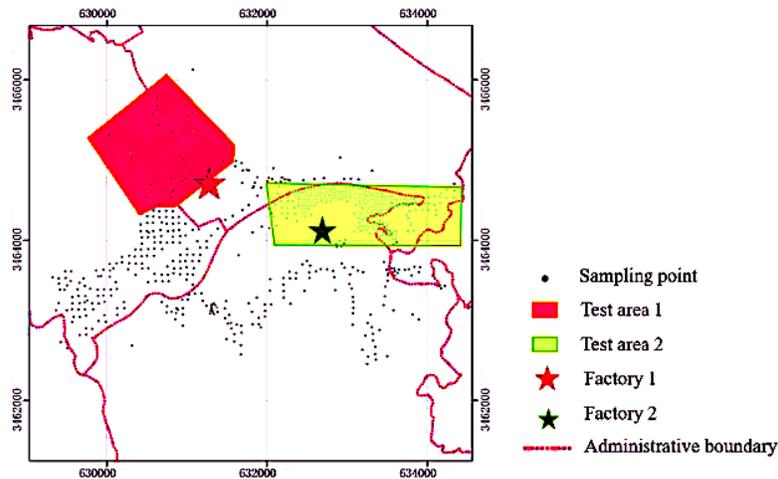


Figure 2. Study area

The primary agricultural products of the area include rice, wheat, rape, and maize. Natural phosphate resources are regionally plentiful, constituting one-sixth of China's phosphate resources. The annual local fertilizer production with phosphorus is around 6M metric tons. The phosphate companies were built in 1958 and have expanded since 2005. The current phosphate capability is 6M tons per year, whereas the overall manufacturing capacity of fertilizer is 7M tons. The two firms depicted in Figure 2 within the examined region are phosphate facilities. Company 1 was founded in 2002, with a yearly manufacturing ability of 4.1M tons. The yearly phosphate production for Company 2 was 500,000 tons. Extensive phosphorus occurring and manufacturing have resulted in significant soil and water contamination, suggesting potential risks to people and the natural world.

1,068 soil specimens were gathered. Sampling sites were established on 90 m periodic grids, further refined to 45 m near two fertilizer facilities. Specimen positions verified using outdoors. Three to five superficial soil specimens were amalgamated at every sampling site to yield a representative aggregation specimen. After eliminating substantial dust and rocks, the collected specimens were sun-dried for seven days and cropped to a particle size of less than 2 mm. Before analysis, the components were preserved in jars inside a heat-regulated condition (5°C).

A pH meter measured the soil pH at a solid-to-fluid proportion of 1:6. Soil As levels were analyzed per the China Standard. Materials have been crushed and filtered to a particle size of less than 0.5 mm. 0.4 g of soil underwent digestion in a mixture comprising 2 ml of hydrofluoric (HF) acids, 4 ml of nitric acids (HNO₃), 2 ml of hydrochloric acids (HCl), and 2 ml of hydrogen peroxides (H₂O₂). The acquired solution was analyzed for contamination, and the soil content was determined. Table presents the amounts of As and the pH levels of the soil. The accepted and empty specimens were established to validate the exactness and precision of the chemical tests.

Feature Extracting Process

A picture of the research location with a resolution of 0.5 meters was acquired. The amount of clouds was 0% when the picture was captured, and a geometrical adjustment was performed using Environment 10.5. The picture was aligned with the geodetic coordinate systems for the sample sites. The picture pixels were allocated relative supervises, and characteristics were obtained utilizing Python and the Geospatial Data Abstraction Framework. Characteristics comprising the highest and lowest band levels, ratios of two groups, and other indexes were computed. Scores from adjacent elements were obtained, and the average standard variations were measured.

The positions of detected elements in the picture (e.g., lakhs, plants, and industries) were designated. The distances and variations among sample sites and labeled elements were computed using the following equations (Equations (1) to (3)):

$$D = \sqrt{(p_x - p_t)^2 + (q_x - q_t)^2} \quad (1)$$

$$G_p = \frac{p_x - p_t}{\min(D)} \quad (2)$$

$$G_q = \frac{q_x - q_t}{\min(D)} \quad (3)$$

Where p_x and q_x denote the dimensions of point x, and p_t and q_t Represents the dimensions of a specified element correspondingly.

Samples taken near the two fertilizer facilities had heightened danger levels, indicating that these industries were significant sources of contamination. Elevated contamination levels were predominantly located northeast of the plants. Detectable inputs of potential pollutant release from the facilities were designated, comprising the buildings, open land, chemical storage sites, and regions. The spacing and variations of sample location to the closest origin were computed.

Forecasting Function

The collected soil specimens were categorized into the entire research region, testing region 1, and testing region 2. All groups randomly designated 50% of the sampling points as the training database, while 50% were allocated for validation. The prediction classifications were taught and developed using the following types of categorization methods: (i) SVM, (ii) MLP, (iii) RD, and (iv) ERF. The final technique is the additional tree classification system, a variant of RF characterized by reduced variance and heightened bias. ERF correlates with enhanced randomness and improved precision in classification. Each model underwent training 500 times with varying random states to provide robust findings. Five hundred forecasting scores were obtained for every designated location, and the median was defined as the forecasted score.

Verification

Modeling forecasts were assessed by analyzing them with validating data sources, using evaluation variables based on risk categorization (i.e., lower, moderate, or high). Kappa scores of 0.3-0.7 signify medium consistency, 0.7-0.9 denote strong consistency and values over 0.8 represent near-perfect consistency. Ordinary and essential kriging interpolating and Inverse Dimension Weighted Interpol (IDWI) were conducted to establish a baseline for comparison with the predictive modeling. The Kriging interpolation technique necessitates that the data adhere to a typical distribution. The contamination content is very skewed. To facilitate Kriging, the Box-Cox conversation was initially performed on the database to approach the usual distribution of the information. Over 40 variable permutations were executed for each sample location, and the mean of the predicted scores obtained using connection approaches was utilized as the resultant forecasting score.

Results and Discussions

The SP levels in database A ranged from 6.12 to 11.42 g/g, with a 1.25 $\mu\text{g/g}$ variation. Database B exhibited ranges from 7.21 to 15.0 ng/g, with a variation of 1.17 $\mu\text{g/g}$. The National Standard of China establishes a contamination risk limit for farming SPs at 20 g/g. The analysis of soil specimens indicated that the SPs in the studied region fell within an acceptable range. The limit for SP danger on farm land is 20 g/g; the SP levels in the two research locations are low, indicating minimal influence on the soil. The research's location is unsuitable for agriculture, as earlier research suggests that SP levels over 10 $\mu\text{g/g}$ might impede agricultural yields.

Feature Selection

The coordinates of the sample locations were utilized to choose the spectrum in the drone-mounted hyperspectral photos. Following an empirical analysis, the spectrum mean of a 3 x 3- vector was utilised as the spectral information. The present research employed cross-validation for characteristics chosen after preprocessing. The best number of repetitions is determined using rounds of Monte Carlo mixed with error. Certain feature bands are picked at specific intervals. According to the findings in Figures 3 and 4, 30 bands

were chosen from database A after 121 repetitions, whereas 21 bands were selected from database B after 139 repetitions. The categories with the lowest error are thus selected as the optimal selections.

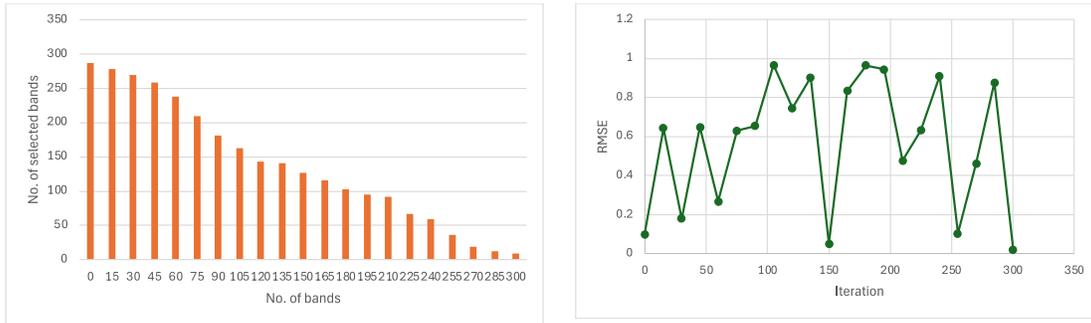


Figure 3. Database A feature selection analysis

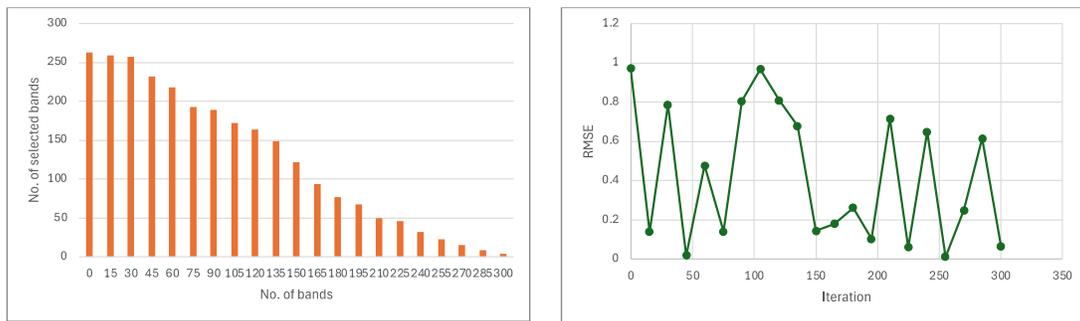


Figure 4. Database B feature selection analysis

Regression Model

The spectrum information was analyzed using MLP, SVM, RF, and Deep Neural Network (DNN) models for regression, with identical models applied to both databases for comparison analysis. Upon successful completion of cross-validation, the model predicts each sample. The outcomes of the four different models over the two databases are as follows. The highest value is 1; the nearer it is to 1, the superior the model's predictive capability. The proximity of the red dot to the 1:1 line in the forecasting algorithm indicates greater precision of the framework. The red dots for every design are closely aligned, with only a few dots positioned further away. The DNN algorithm has superior forecasting accuracy. Among the several models, the red dot representing the DNN algorithm is nearest to the 1:1 line, indicating a nearly perfect match.

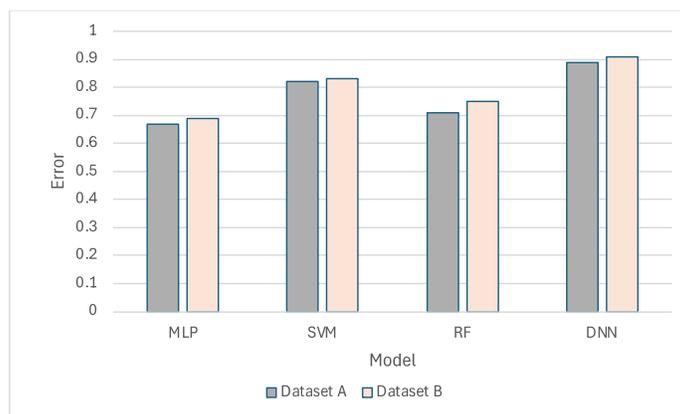


Figure 5. Evaluation analysis

Figure 5 indicates that the predictive outcomes of different methods exhibit great accuracy. The system employed in prior research has an error of 0.75. In this investigation, the error value of the modeling in both databases surpasses 0.75; errors arise when forecasting significant values, leading to an elevated value. The SVM, a widely utilized model for soil spectra forecasting, has significant consistency and achieves good accuracy in predicting for both databases. The error of the handheld spectrometry used in prior investigations is around 0.8. The RF model needs more samples to leverage its benefits in scenarios with few samples. The findings align with those of previous investigations employing hand-held spectrometry. The metrics for the DNN system on the two databases exhibit the highest accuracy. The results indicate that the DNN forecasting system is superior to the four assessment models across the two databases. The quality of the soil specimens will limit the model's estimating capability. If the soil surroundings have values over 20 g/g, the precision of the model requires more investigation.

Many devices are employed in the methodologies for predicting SPs. The predominant technique is the predictive method with hand-held spectrometry. This used an outfitted with a hyperspectral camera to gather spectrum information. The forecast accuracy is comparable to the hand-held spectrometer, serving as a viable alternative.

Overall SP Distribution Map

The assessment findings indicate that the DNN exhibits the highest result. The DNN is employed to produce the SP dispersion mapping. The research categorized the anticipated values into five ranges based on the group's mean value. Five periods are designated by distinct colors: dark green (0-7 $\mu\text{g/g}$), light-colored green (7-12 $\mu\text{g/g}$), yellow (12-15 $\mu\text{g/g}$), orange (15-20 $\mu\text{g/g}$), and red (20+ $\mu\text{g/g}$). The resulting SP dispersion map for database A indicates that the SP concentration in the SP is below eight $\mu\text{g/g}$, with the red areas (exceeding 14 $\mu\text{g/g}$) being comparatively uncommon. This aligns with the average of the sample measuring findings. The SP concentration in the probably polluted mining region is relatively low. The vehicle's passage resulted in an elevated level of SP due to soil compaction and exposure.

Database B is a potentially polluted location, exhibiting indications of probable dry riverbeds on the surface. The deposition of the slag is considered to have occurred due to precipitation. It indicates that SP continues to enter the flowing water and requires remediation. The mean of database B (average 10.54 $\mu\text{g/g}$) is much greater than database A (average 10.24 $\mu\text{g/g}$). The SP dispersion mapping indicates that the quantity of red segments (exceeding 15 $\mu\text{g/g}$) in Database B is markedly more significant than in Database A.

The two SP dispersion maps databases demonstrate that employing a DNN model with UAV hyperspectral imagery is an excellent way to generate SP dispersion diagrams, providing substantial assistance for soil administration and surveillance. Both databases are 2 kilometers apart and were obtained on different days. It indicates a specific level of universality.

Extract the Spectrum

In this investigation, the drone operated at an altitude of 50 m, with hyperspectral pictures exhibiting a spatial resolution of around 0.04 m. The spectral means of 1 x 1, 3 x 3, and 5 x 5- vectors were employed for analysis to ascertain the appropriate size. The following evaluation was used to identify bands incorporated into the predictive model through extensive experimentation. The research utilized the spectral means of various pixel dimensions as input for the predictive model. The hyperspectral combination of different methods demonstrates an overall result with an error above 0.7 in database A.

Obtaining hyperspectral information from pictures exhibits instability, with the error being more minor on database B. Following the standard treatment of 3 x 3 pixels, both databases' accuracy is enhanced, suggesting that the spectrum mean can augment durability. A spectrometer is employed to acquire the spectra of an SP specimen, combining the obtained data on several occasions. Using the 5 x 5 pixel mean for filtering diminishes the model's precision due to the spectral difference of the soil, as the averaged spectrum encompasses several spectra that are entirely unlike. The assessment findings indicate that the DNN framework possesses distinct benefits compared to the other methods in figure 6 and 7.



Figure 6. Database A dispersion analysis

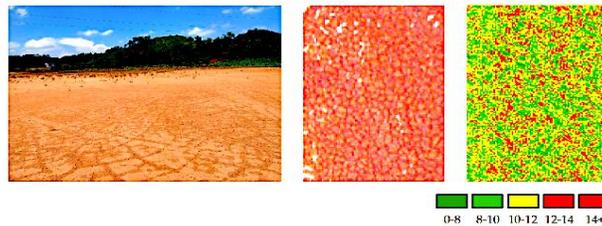


Figure 7. Database B dispersion analysis

The height at which drones gather flight data is significantly correlated with the median size of the spectral frame. The typical spectrum window size employed in several research is 3 * 3 when the height of the drone-mounted camera is below 100 meters. When the flight height of a drone exceeds 100 meters, most studies employ a single pixel. Spectrum unmixing enhances precision in investigations of altitudes beyond 1000 meters. The geographic unity of the predicted materials with the algorithm's forecast is significant. The constituents of soil, water, and vegetation are distinctly dissimilar, and the spectrum varies significantly. Several trials are necessary to determine the suitable processing procedure for the spectrum extraction process, considering the varying flight heights and assessment elements. Suitable methodologies can enhance the predictive accuracy of the algorithm.

Conclusion

This research showed that the precision and generalizability of ML algorithms for predicting SP at locations might be enhanced by selecting highly correlated indices and minimizing redundant indices using feature fusion. The interpreting study of the models indicated that the index of fundamental data, contamination potential from products and raw materials, and the SP control level of companies varied in their significance regarding the impact on SP at sites. The stratum data indices showed no significant relevance to categorizing SP identification at the locations. In the ecological oversight of PCS, greater emphasis should be placed on aspects associated with polluting sources and routes and the amount of pollution management. Certain indices and their evaluation criteria within the technical specifications for assessing SP and risk should be refined, enhancing the precision of identifying outcomes. Future studies will leverage big data and advanced models to enhance the effectiveness and uniformity of site questionnaire gathering, thereby addressing challenges in

acquiring site ecological data and augmenting the accuracy and generalizability of ML techniques in detecting SP from PCS.

Author Contributions

All Authors contributed equally.

Conflict of Interest

The authors declared that no conflict of interest.

References

- Angin, P., Anisi, M.H., Göksel, F., Gürsoy, C., & Büyükgülçü, A. (2020). Agrilora: a digital twin framework for smart agriculture. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 11(4), 77-96.
- Askari, M. S., Alamdari, P., Chahardoli, S., & Afshari, A. (2020). Quantification of heavy metal pollution for environmental assessment of soil condition. *Environmental Monitoring and Assessment*, 192, 1-17.
- Baier, C., Modersohn, A., Jalowy, F., Glaser, B., & Gross, A. (2022). Effects of recultivation on soil organic carbon sequestration in abandoned coal mining sites: a meta-analysis. *Scientific Reports*, 12(1), 20090. <https://doi.org/10.1038/s41598-022-22937-z>
- Camgözlü, Y., & Kutlu, Y. (2023). Leaf Image Classification Based on Pre-trained Convolutional Neural Network Models. *Natural and Engineering Sciences*, 8(3), 214-232.
- Danková, Z., Štyriaková, I., Kovaničová, L., Čechovská, K., Košuth, M., Šuba, J., Nováková, J., Konečný, P., Tuček, L., Žecová, K., Lenhardtová, E., & Németh, Z. (2021). Chemical Leaching of Contaminated Soil – Case Study. *Archives for Technical Sciences*, 1(24), 65–72.
- Deiss, L., Margenot, A. J., Culman, S. W., & Demyan, M. S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, 365, 114227. <https://doi.org/10.1016/j.geoderma.2020.114227>
- Elizondo-Martinez, E. J., Andres-Valeri, V. C., Rodriguez-Hernandez, J., & Sangiorgi, C. (2020). Selection of additives and fibers for improving the mechanical and safety properties of porous concrete pavements through multi-criteria decision-making analysis. *Sustainability*, 12(6), 2392. <https://doi.org/10.3390/su12062392>
- Hu, B., Shao, S., Ni, H., Fu, Z., Huang, M., Chen, Q., & Shi, Z. (2021). Assessment of potentially toxic element pollution in soils and related health risks in 271 cities across China. *Environmental Pollution*, 270, 116196. <https://doi.org/10.1016/j.envpol.2020.116196>
- Huang, Y., Harilal, S. S., Bais, A., & Hussein, A. E. (2023). Progress toward machine learning methodologies for laser-induced breakdown spectroscopy with an emphasis on soil analysis. *IEEE Transactions on Plasma Science*, 51(7), 1729-1749.

- Jas, K., & Dodagoudar, G. R. (2023). Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dynamics and Earthquake Engineering*, 165, 107662. <https://doi.org/10.1016/j.soildyn.2022.107662>
- Khan, S., Naushad, M., Lima, E. C., Zhang, S., Shaheen, S. M., & Rinklebe, J. (2021). Global soil pollution by toxic elements: Current status and future perspectives on the risk assessment and remediation strategies—A review. *Journal of Hazardous Materials*, 417, 126039. <https://doi.org/10.1016/j.jhazmat.2021.126039>
- Li, K., & Sun, R. (2024). Understanding the driving mechanisms of site contamination in China through a data-driven approach. *Environmental Pollution*, 342, 123105. <https://doi.org/10.1016/j.envpol.2023.123105>
- Li, X., Wu, Y., Leng, Y., Xiu, D., Pei, N., Li, S., & Tian, Y. (2023). Risk assessment, spatial distribution, and source identification of heavy metals in surface soils in Zhijin County, Guizhou Province, China. *Environmental Monitoring and Assessment*, 195(1), 132. <https://doi.org/10.1007/s10661-022-10674-9>
- Maurya, S., Abraham, J. S., Somasundaram, S., Toteja, R., Gupta, R., & Makhija, S. (2020). Indicators for assessment of soil quality: a mini-review. *Environmental Monitoring and Assessment*, 192, 1-22.
- Mohamed, S., Kumaran, U., & Rakesh, N. (2024). An Approach towards Forecasting Time Series Air Pollution Data Using LSTM-based Auto-Encoders. *Journal of Internet Services and Information Security*, 14(2), 32-46.
- Mosavi, A., Samadianfard, S., Darbandi, S., Nabipour, N., Qasem, S. N., Salwana, E., & Band, S. S. (2021). Predicting soil electrical conductivity using multi-layer perceptron integrated with grey wolf optimizer. *Journal of Geochemical Exploration*, 220, 106639. <https://doi.org/10.1016/j.gexplo.2020.106639>
- Obiri-Nyarko, F., Duah, A. A., Karikari, A. Y., Agyekum, W. A., Manu, E., & Tagoe, R. (2021). Assessment of heavy metal contamination in soils at the Kpone landfill site, Ghana: Implication for ecological and health risk assessment. *Chemosphere*, 282, 131007. <https://doi.org/10.1016/j.chemosphere.2021.131007>
- Paul, P. K., Sinha, R. R., Aithal, P. S., Aremu, B., & Saavedra, R. (2020). Agricultural Informatics: An Overview of Integration of Agricultural Sciences and Information Science. *Indian Journal of Information Sources and Services*, 10(1), 48–55.
- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), 2495-2527
- Ray, A., Kumar, V., Kumar, A., Rai, R., Khandelwal, M., & Singh, T. N. (2020). Stability prediction of Himalayan residual soil slope using artificial neural network. *Natural Hazards*, 103(3), 3523-3540.
- Van Der Westhuizen, S., Heuvelink, G. B., & Hofmeyr, D. P. (2023). Multivariate random forest for digital soil mapping. *Geoderma*, 431, 116365. <https://doi.org/10.1016/j.geoderma.2023.116365>

- Vinante, C., Sacco, P., Orzes, G., & Borgianni, Y. (2021). Circular economy metrics: Literature review and company-level classification framework. *Journal of Cleaner Production*, 288, 125090. <https://doi.org/10.1016/j.jclepro.2020.125090>
- Wood, S. A., & Blankinship, J. C. (2022). Making soil health science practical: guiding research for agronomic and environmental benefits. *Soil Biology and Biochemistry*, 172, 108776. <https://doi.org/10.1016/j.soilbio.2022.108776>
- Ye, M., Zhu, L., Li, X., Ke, Y., Huang, Y., Chen, B., & Feng, H. (2023). Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Science of The Total Environment*, 858, 159798. <https://doi.org/10.1016/j.scitotenv.2022.159798>
- Zhai, H., Lv, C., Liu, W., Yang, C., Fan, D., Wang, Z., & Guan, Q. (2021). Understanding spatio-temporal patterns of land use/land cover change under urbanization in Wuhan, China, 2000–2019. *Remote Sensing*, 13(16), 3331. <https://doi.org/10.3390/rs13163331>