

# RED TEAMING MODELS FOR HYPERSPECTRAL IMAGE ANALYSIS USING EXPLAINABLE AI

Vladimir Zaigrajew<sup>1</sup>   Hubert Baniecki<sup>1,2</sup>   Lukasz Tulczyjew<sup>3,4</sup>   Agata M. Wijata<sup>3,4</sup>

Jakub Nalepa<sup>3,4</sup>

Nicolas Longépé<sup>5</sup>

Przemyslaw Biecek<sup>1,2</sup>

<sup>1</sup>Warsaw University of Technology, <sup>2</sup>University of Warsaw, <sup>3</sup>Silesian University of Technology,  
<sup>4</sup>KP Labs, <sup>5</sup>European Space Agency

<sup>1</sup>vladimir.zaigrajew.dokt@pw.edu.pl

## ABSTRACT

Remote sensing (RS) applications in the space domain demand machine learning (ML) models that are reliable, robust, and quality-assured, making red teaming a vital approach for identifying and exposing potential flaws and biases. Since both fields advance independently, there is a notable gap in integrating red teaming strategies into RS. This paper introduces a methodology for examining ML models operating on hyperspectral images within the HYPERVIEW challenge, focusing on soil parameters’ estimation. We use post-hoc explanation methods from the Explainable AI (XAI) domain to critically assess the best performing model that won the HYPERVIEW challenge and served as an inspiration for the model deployed on board the INTUITION-1 hyperspectral mission. Our approach effectively red teams the model by pinpointing and validating key shortcomings, constructing a model that achieves comparable performance using just 1% of the input features and a mere up to 5% performance loss. Additionally, we propose a novel way of visualizing explanations that integrate domain-specific information about hyperspectral bands (wavelengths) and data transformations to better suit interpreting models for hyperspectral image analysis.

## 1 INTRODUCTION

Remotely-sensed hyperspectral imaging presents significant opportunities in Earth observation, offering an extensive amount of data and potential for global scalability (Wijata et al., 2023). Hyperspectral images (HSIs) capture numerous spectral bands, providing detailed insights about scanned objects. In precision agriculture, HSIs are crucial for extracting various soil parameters, enabling non-invasive, large-scale monitoring to enhance agricultural practices (Nalepa et al., 2022). However, deploying machine learning (ML) models for these tasks faces challenges in creating diverse, high-quality datasets, often constrained by time, cost, and human error. The HYPERVIEW: “Seeing beyond the visible” challenge, with its carefully curated dataset, aimed to address these issues by developing models to retrieve soil parameters from HSIs. Attracting substantial participation (160 registered teams, with ca. 50 actively participating) from the research community (Nalepa et al., 2022), the challenge resulted in a high-performing ML model. The winning solution was set for deployment on the INTUITION-1 satellite. However, before its deployment on the INTUITION-1 satellite, the winning challenge solution had to undergo a thorough performance evaluation.

This important deployment phase reflects the wider industry trend of continuous model improvement, where the use of various *red teaming* strategies is essential. These strategies aim to uncover flaws and biases across diverse model families such as large language models (Perez et al., 2022; Ganguli et al., 2022), text-to-image models (Mehrabi et al., 2023), and diffusion models (Rando et al., 2022). In parallel, stakeholders in remote sensing (RS) call for higher explainability standards (Roscher et al., 2020a; Gevaert, 2022) as explainable AI (XAI) methods have shown to be useful in many other practical applications (Roscher et al., 2020b). Therefore, several works (Singh et al., 2022; Turan et al., 2023; Dantas et al., 2023; Emam et al., 2023b; Ekim & Schmitt, 2023;

[Abbas et al., 2023] have emerged to explain models used in RS. However, only a few studies have shown how applying XAI can pinpoint performance gaps in models [Emam et al., 2023a] or improve predictions [De Lucia et al., 2022], effectively aligning XAI with *red teaming* strategies. Our work aims to advance this integration, focusing specifically on models from the HYPERVIEW challenge designed for estimating soil parameters. We adapt post-hoc explanations to the specific domain and use existing knowledge about the model to generate more sophisticated explanatory visualizations, which are used to successfully red team the model.

**Contribution.** Our key contributions are: (1) We demonstrate how, by using SHapley Additive exPlanations (SHAP), we were able to conduct an in-depth analysis of the EAGLEEYES model [Kuzu et al., 2022] within the HYPERVIEW challenge, laying the groundwork for unraveling shortcomings in model performance. (2) We present a novel way of visualizing explanations that integrate domain-specific information about hyperspectral bands and data transformations, aimed at boosting the red teaming of hyperspectral image analysis models. (3) The result, aimed at red teaming the model, was the development of a model pruning technique focused on feature selection. This method, based on SHAP, yielded a more efficient model without compromising its regression performance.

## 2 METHODOLOGY

### 2.1 SHAPLEY VALUES AS A MODEL AGNOSTIC TOOL FOR KNOWLEDGE EXTRACTION

Shapley values for predictive models [Robnik-Šikonja & Kononenko, 2008; Lundberg & Lee, 2017] are one of the most popular XAI methods [Holzinger et al., 2022]. This well-deserved popularity is due to such properties as model agnosticity (they can be calculated for any model structure), additivity (so that the contribution of a group of features is the sum of the contributions of the features in that group), the availability of efficient implementations (e.g., TreeSHAP for tree-based models [Lundberg et al., 2020]), and the possibility of use in the process of model analysis [Biecek & Burzykowski, 2021]. For a detailed mathematical explanation, please refer to Appendix A.2. Shapley values are initially computed locally for individual features and observations, but they can also be aggregated to get the global importance of a selected variable, or attribution of a group of features. In the following sections, we will aggregate features corresponding to the same spectral band or the same data transformation method.

### 2.2 DATA AND MODELS

We work with the HYPERVIEW dataset and the winning HYPERVIEW model EAGLEEYES. The dataset, acquired on March 3, 2021, using the HySpex VS-725 (Norsk Elektro Optikk AS) acquisition system, consists of images from two SWIR-384 imagers and one VNIR-1800 imager aboard the Piper PA-31 Navajo aircraft. After undergoing comprehensive preprocessing and calibration as detailed by [Nalepa et al., 2022], the images were aligned with in-situ soil parameter measurements taken using the Mehlich 3 methodology. From each parcel, 12 soil samples were collected and analyzed in the laboratory. This process provided single, weakly-labeled, per-image ground-truth data for four key soil parameters: *phosphorus* ( $P$ ), *potassium*, ( $K$ ) *magnesium* ( $Mg$ ) and soil acidity ( $pH$ ). Our study analyzes two dataset versions: HYPERVIEW with 150 bands (462–942 nm) and INTUITION-1 with 192 bands (460–920 nm), matching the spectral characteristics of a hyperspectral camera mounted on board Intuition-1, a 6U satellite (KP Labs, Poland). The EAGLEEYES model, available at [https://github.com/ridvansalihkuzu/hyperview\\_eagleeyes](https://github.com/ridvansalihkuzu/hyperview_eagleeyes), was provided with and without spatial features extracted and deployed over two aforementioned versions of the dataset. For detailed information on the dataset and model, refer to Appendix A.1.

### 2.3 MODEL PRUNING

ML models, particularly complex ones, require substantial computational resources for training and inference. This is challenging for real-time processing and deployment on resource-constrained devices like imaging satellites. Model pruning, a technique to reduce model size without affecting performance, addresses this issue. It is widely used in both classic ML [Zhou & Mentch, 2023] and deep learning [Cheng et al., 2023] to optimize models for specific tasks, including those in RS [Qi et al., 2019; Guo et al., 2021; Zhou et al., 2018]. However, traditional pruning methods often depend

on time-consuming processes and can introduce constraints. Our approach utilizes explanations derived from Shapley values for effective feature selection in model pruning. This method significantly reduces the model’s size and computational demands while preserving its performance, resulting in lighter and faster models suitable for edge devices.

### 3 EXPLANATORY MODEL ANALYSIS

In this section, we conduct a thorough analysis of the HYPERVIEW challenge, focusing on evaluating the top-ranked model’s performance through Shapley values. Our goal is to identify and highlight any shortcomings in these models, offering insights for enhancements in their ongoing deployment. The initial analysis of the models and datasets is detailed in Appendix A.1. A critical finding was the poor model performance, notably in consistently predicting 90% of values within a narrow range, as seen in Figure 4. This led to significant mispredictions, especially for outliers in each soil parameter.

**SHAP** Our initial Shapley values analysis examined feature importance distribution across soil parameters and model versions, with detailed visualizations in Appendix A.4. Figure 5 displays the top features of the EAGLEEYES model and their importance in predicting each soil parameter. Meanwhile, Figure 7 provides beeswarm plots for each version of the EAGLEEYES model in predicting *phosphorus*, highlighting the top features and their impact on predictions as defined by Shapley values. Key findings include the models relying on a limited set of features (less than 1% of the total), likely causing their narrow prediction range. Each model version favors different features, and as the number of features used increases, the focus shifts from the most influential ones, hinting at possible **underfitting** and reinforcing the limited prediction range observed.

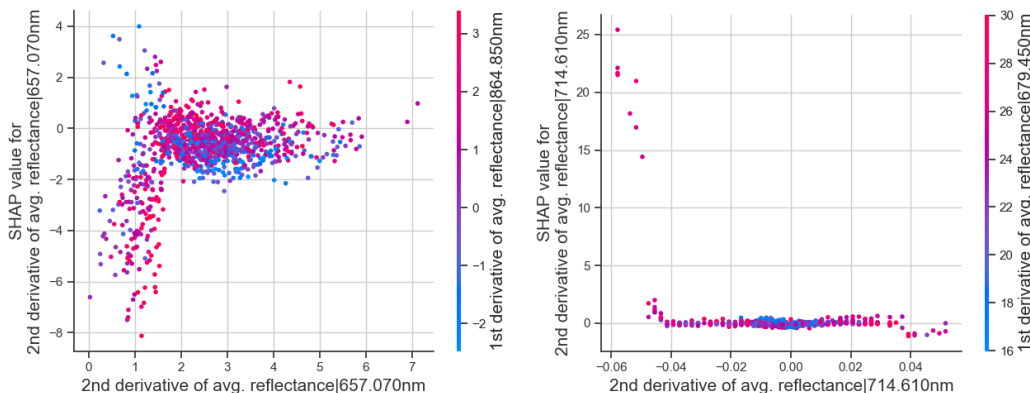


Figure 1: Variations in Shapley values in response to changing feature values in the EAGLEEYES model for *phosphorus* predictions, highlighting the most influential feature (**left**) and a feature with lesser impact (**right**).

While popular feature importance methods provide similar insights, Shapley values offer a more detailed analysis of how individual features and specific samples impact model predictions. Analyzing SHAP’s dependency and beeswarm plots, we gain an understanding of the effects of individual feature value fluctuations on predictions. Figure 1 illustrates this in two parts: the left plot demonstrates how Shapley values correlate with the most impactful feature in the EAGLEEYES model, peaking at a value of 2 and then stabilizing, suggesting a limited utilization of feature values for even most impactful features. The right plot, targeting a less influential feature, indicates the model’s bias towards memorizing specific outliers, resulting in prediction variability around the established prediction window. These observations support our hypothesis regarding the model’s constrained prediction range, which primarily depends on key features for **simplistic** value estimation and incorporates **outlier memorization** for adaptability within its prediction range.

**Aggregation Analysis** Shapley values, formatted as `<n_samples, features, class>`, enable aggregation into **hyperspectral bands** and **data transformation** groups. Figure 2 demonstrates the significance of features, grouped by preprocessing transformations on the  $y$ -axis and

wavelength (hyperspectral band) on the  $x$ -axis. This visualization aids in understanding model interpretations within the spectral domain of HSIs, incorporating preprocessing transformations. Notably, this plot reveals that key features are distributed across various bands instead of being clustered in specific areas. Aggregated transformation visualizations, detailed in Appendix A.5 offer more informative insights. Figure 8 highlights that spectral gradient features significantly influence the EAGLEYES model’s predictions, unlike other transformations with minimal impact. Additionally, models incorporating spatial data tend to rely more on spatial features for predictions, underscoring their critical role in enhancing model performance.

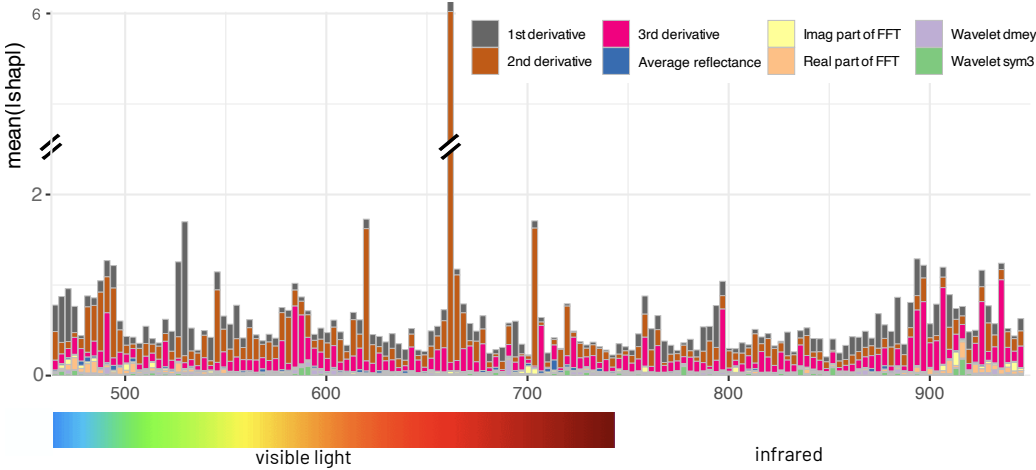


Figure 2: Detailed visualization of Shapley values, depicting the significance of various aggregated feature transformation groups with their corresponding wavelengths aggregations on the  $x$ -axis.

**Model Pruning** In our concluding analysis, we assessed model performance using key features identified through Shapley value analysis. Beeswarm and feature importance plots enabled the efficient selection of a feature subset for the prediction of soil parameters. Table 1 shows that models with a reduced number of input features reached Mean Absolute Error (MAE) scores comparable to those using all features for each soil parameter. Models trained with spatial features presented challenges in selecting feature subsets for training. It was expected due to smoother feature importance distribution. Nonetheless, pruned models with under 1% of features performed similarly to baseline models, as a two-sample t-test confirmed, with minimal metric decline. This affirms our hypothesis about the models’ dependency on a limited set of features for precise predictions.

Dataset	F. selection	$P$	$K$	$Mg$	$pH$
HYPERVIEW	✗ (1200)	22.6	48.4	31.3	0.206
	✓ (3)	23.3 (+3%)	48.7 (+1%)	32.6 (+4%)	0.213 (+4%)
INTUITION-1	✗ (1490)	22.5	48.5	30.9	0.204
	✓ (3)	22.7 (+1%)	48.5 (0%)	32.5 (+5%)	0.214 (+5%)
HYPERVIEW (spatial)	✗ (2400)	22.4	47.8	31.0	0.205
	✓ (5)	23.0 (+3%)	48.9 (+2%)	32.5 (+5%)	0.212 (+3%)
INTUITION-1 (spatial)	✗ (3026)	22.3	47.8	31.1	0.204
	✓ (6)	23.2 (+4%)	50.1 (+5%)	32.4 (+4%)	0.209 (+2%)

Table 1: Comparison of model performance on the test dataset using MAE ( $\downarrow$ ) for soil parameters ( $P$ ,  $K$ ,  $Mg$ ,  $pH$ ). Evaluations cover two scenarios: trained on original features (✗) and trained on post-feature selection (✓), noting the feature count, e.g., 3 of 1200 for HYPERVIEW .

## 4 CONCLUSION

We developed an innovative red teaming methodology for hyperspectral image analysis models, focusing on SHAP to utilize explainability and apply assurance of robustness. Our method, employed on the winning HYPERVIEW challenge model (EAGLEEYES), uncovers and visualizes model failures and biases, offering advanced visualizations into hyperspectral bands and data transformation aggregations. Notably, we found the EAGLEEYES model utilized less than 1% of available features, leading us to create simpler but effective model alternatives. Our study marks a significant advance in the application of red teaming using explainable AI to the hyperspectral imaging domain models.

## ACKNOWLEDGEMENTS

We thank Bogdan Ruszczak for his valuable feedback about this work. The work on this project is financially supported by the European Space Agency grant (ESA AO/1-11524/22/I-DT) and it was supported by  $\Phi$ -lab (<https://philab.esa.int/>). Lukasz Tulczyjew, Agata M. Wijata and Jakub Nalepa were supported by the Silesian University of Technology through the grant for maintaining and developing research potential.

## REFERENCES

- Adel Abbas, Michele Linardi, Etienne Vareille, Vassillis Christophides, and Claudia Paris. Towards explainable ai4eo: An explainable deep learning approach for crop type mapping using satellite images time series. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1088–1091. IEEE, 2023.
- Przemyslaw Biecek and Tomasz Burzykowski. *Explanatory Model Analysis*. Chapman and Hall/CRC, New York, 2021.
- Hongrong Cheng, Miao Zhang, and Javen Qinfeng Shi. A survey on deep neural network pruning-taxonomy, comparison, analysis, and recommendations. *arXiv preprint arXiv:2308.06767*, 2023.
- Ian Covert, Scott M Lundberg, and Su-In Lee. Understanding Global Feature Contributions With Additive Importance Measures. In *NeurIPS*, 2020.
- Cassio F. Dantas, Thalita F. Drumond, Diego Marcos, and Dino Ienco. Counterfactual Explanations for Remote Sensing Time Series Data: An Application to Land Cover Classification. In *ECML PKDD*, 2023.
- Gianluca De Lucia, Marco Lapegna, and Diego Romano. Towards explainable ai for hyperspectral image classification in edge computing environments. *Computers and Electrical Engineering*, 103:108381, 2022.
- Burak Ekim and Michael Schmitt. Explaining Multimodal Data Fusion: Occlusion Analysis for Wilderness Mapping. In *ICLR ML4RS Workshop*, 2023.
- Ahmed Emam, Ribana Roscher, and Mohamed Farag. Confident naturalness explanation (cne): A framework to explain and assess patterns forming naturalness in fennoscandia with confidence. In *Northern Lights Deep Learning Conference*, number FZJ-2024-00529. Pflanzenwissenschaften, 2023a.
- Ahmed Emam, Timo T Stomberg, and Ribana Roscher. Leveraging activation maximization and generative adversarial training to recognize and explain patterns in natural areas in satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 2023b.
- Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, et al. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. *arXiv preprint arXiv:2209.07858*, 2022.
- Caroline M Gevaert. Explainable ai for earth observation: A review including societal and regulatory perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 112: 102869, 2022.

- Xianpeng Guo, Biao Hou, Bo Ren, Zhongle Ren, and Licheng Jiao. Network pruning for remote sensing images classification based on interpretable cnns. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.
- Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. Explainable AI Methods – A Brief Overview. In *xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020*, pp. 13–38, 2022.
- Rıdvan Salih Kuzu, Frauke Albrecht, Caroline Arnold, Roshni Kamath, and Kai Konen. Predicting soil properties from hyperspectral satellite images. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4296–4300. IEEE, 2022.
- Scott M. Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *NeurIPS*, 2017.
- Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.
- Ninareh Mehrabi, Palash Goyal, Christophe Dupuy, Qian Hu, Shalini Ghosh, Richard Zemel, Kai-Wei Chang, Aram Galstyan, and Rahul Gupta. Flirt: Feedback loop in-context red teaming. *arXiv preprint arXiv:2308.04265*, 2023.
- Jakub Nalepa, Bertrand Le Saux, Nicolas Longép e, Lukasz Tulczyjew, Michal Myller, Michal Kawulok, Krzysztof Smykala, and Michal Gumiela. The hyperview challenge: Estimating soil parameters from hyperspectral images. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4268–4272. IEEE, 2022.
- Ethan Perez, Saffron Huang, Francis Song, Trevor Cai, Roman Ring, John Aslanides, Amelia Glaese, Nat McAleese, and Geoffrey Irving. Red Teaming Language Models with Language Models. In *EMNLP*, 2022.
- Baogui Qi, He Chen, Yin Zhuang, Shaorong Liu, and Liang Chen. A network pruning method for remote sensing image scene classification. In *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pp. 1–4. IEEE, 2019.
- Javier Rando, Daniel Paleka, David Lindner, Lennart Heim, and Florian Tram er. Red-Teaming the Stable Diffusion Safety Filter. In *NeurIPS ML Safety Workshop*, 2022.
- Marko Robnik-Šikonja and Igor Kononenko. Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, 20(5):589–600, 2008.
- R Roscher, B Bohn, MF Duarte, and J Garcke. Explain it to me–facing remote sensing challenges in the bio-and geosciences with explainable machine learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:817–824, 2020a.
- Ribana Roscher, Bastian Bohn, Marco F Duarte, and Jochen Garcke. Explainable machine learning for scientific insights and discoveries. *Ieee Access*, 8:42200–42216, 2020b.
- Harpinder Singh, Ajay Roy, Raj Setia, and Brijendra Pateriya. Estimation of chlorophyll, macronutrients and water content in maize from hyperspectral data using machine learning and explainable artificial intelligence techniques. *Remote Sensing Letters*, 13(10):969–979, 2022.
- DE Turan, E Aptoula, A Ert urk, and G Taskin. Interpreting hyperspectral remote sensing image classification methods via explainable artificial intelligence. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5950–5953. IEEE, 2023.
- Agata M Wijata, Michel-Fran ois Foulon, Yves Bobichon, Raffaele Vitulli, Marco Celesti, Roberto Camarero, Gianluigi Di Cosimo, Ferran Gascon, Nicolas Long ep e, Jens Nieke, et al. Taking artificial intelligence into space through objective selection of hyperspectral earth observation applications: To bring the “brain” close to the “eyes” of satellite missions. *IEEE Geoscience and Remote Sensing Magazine*, 11(2):10–39, 2023.

Siyu Zhou and Lucas Mentch. Trees, forests, chickens, and eggs: when and why to prune trees in a random forest. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 16(1): 45–64, 2023.

Yi Zhou, Rui Zhang, Shixin Wang, and Futao Wang. Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy. *Sensors*, 18(7):2013, 2018.

## A APPENDIX

### A.1 DATASET AND MODEL ANALYSIS

**Dataset Analysis** As outlined in Section 2.2, the HYPERVIEW challenge comprises **1732** training and **1154** testing patches with a spectral range of 462.080 – 938.370 **nm** across 150 bands. The dataset’s heterogeneity is evident in varying patch sizes and aspect ratios, challenging the selection of an architectural approach. Notably, approximately 38% of the dataset consists of small patches (up to  $32 \times 32$  pixels), further complicating the analysis. The INTUITION-1 dataset, while spatially identical to HYPERVIEW challenge, offers a distinct spectral perspective (192 bands), providing a broader spectral viewpoint without adding spatial complexity. Analysis of ground truth data reveals a concentrated distribution of soil parameters, with outliers constituting less than 5% of the data. Excluding *pH*, outliers tend to skew right of the mean, while *pH* outliers show more symmetry.

**Model Analysis** The superior EAGLEEYES model from the HYPERVIEW challenge builds upon classic ML models with the combination of feature engineering for supervised regression tasks (Figure 3). This model was developed in four variations, with and without spatial feature extraction, tailored for HSIs of 150 (HYPERVIEW dataset) and 192 (INTUITION-1 dataset) spectral bands. We specify the inclusion of spatial features by adding “spatial” in parentheses when referencing a model (e.g., HYPERVIEW (spatial) benefits from spatial features across 150 bands, while INTUITION-1 focuses solely on spectral features over 192 bands). Performance evaluation involved juxtaposing predicted values against actual ground truth data, and analyzing residuals to assess accuracy. Figure 4 showcases the EAGLEEYES model’s iterations across soil parameters. Despite uniform performance across parameters, the model’s accuracy was far from ideal. Residual analysis, through boxplots and histograms, indicated a Gaussian distribution with a mean of 0, but notable variance points to significant deviations for many samples. Scatter plots in Figure 4 reflect this, showing predictions primarily near the zero-error line but with widespread variation. This distribution leads to higher residuals, especially for values outside the model’s learned prediction range, highlighting challenges with distinct or outlier data.

### A.2 MATHEMATICS BEHIND SHAP

The Shapley value  $\phi_i(x)$  for feature  $i$  and observation  $x$  are defined as

$$\phi_i(x) = \sum_{S \subseteq P \setminus \{i\}} \frac{|S|! (p - |S| - 1)!}{p!} (v_{S \cup \{i\}}(x) - v_S(x)),$$

where  $P$  is a set of all features,  $p = |P|$  is the number of features,  $v_S(x)$  is the prediction calculated for model  $f$  after features from set  $S$  are set, typically it is the mean prediction for the conditional distribution with fixed values of the features in the set  $S$ , i.e.,  $v_S(x) = E_{x|S} f(x)$ , but there are also other alternatives used if features are correlated (Covert et al., 2020). Shapley values are calculated locally for a specific feature and observation, but they can also be aggregated to get the global importance of a selected variable  $imp(i) = \sum_{j=1}^N |\phi_i(x_j)|$ , or attribution of a group of features ( $F \subseteq P$ )  $\varphi_F(x) = \sum_{i \in F} \phi_i(x)$ . In our case, we make use of this ability to aggregate shapley values according to hyperspectral bands or specific transformation methods.

### A.3 SHAP IN RESIDUALS

In our extended analysis, we closely examined instances of significant overestimation and underestimation, as well as the “best” prediction outcomes, across all model variants. We consistently observed the same set of features appearing in each scenario, although their importance rankings varied. This pattern makes it challenging to identify particular features responsible for overestimation and underestimation directly. Moreover, investigating specific prediction cases did not yield insights that could be generalized to understand the model’s overall behavior or the characteristics of the underlying data more broadly.



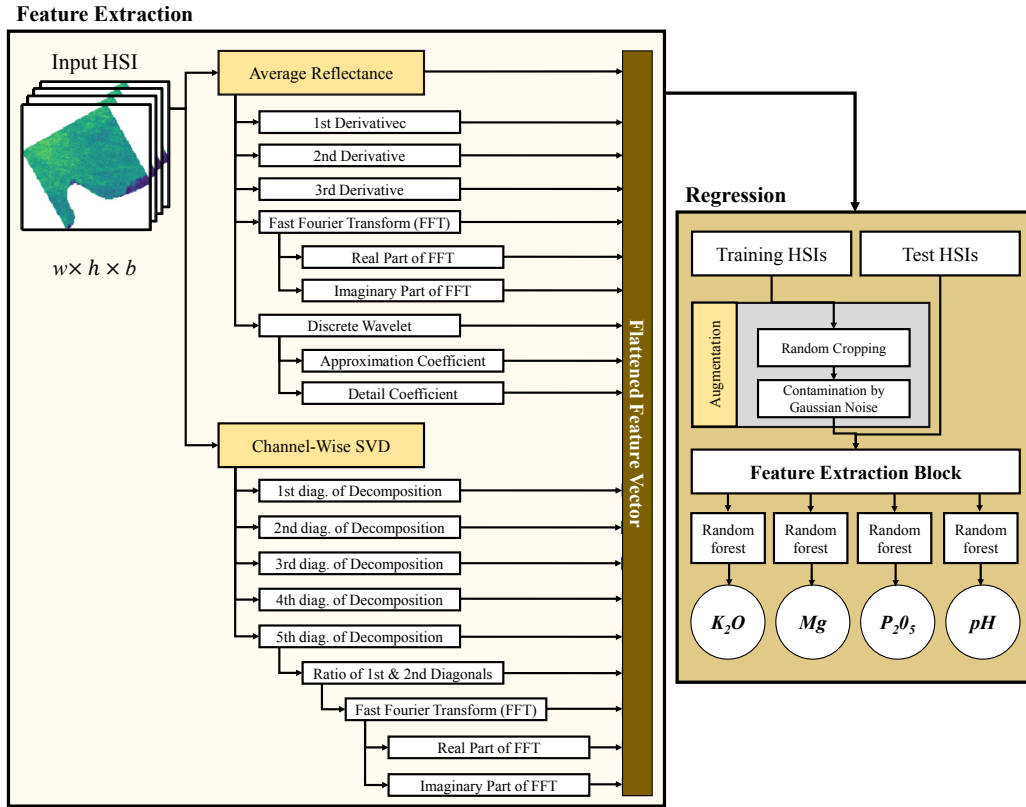


Figure 3: A flowchart presenting the EAGLEEYES soil parameters' estimation approach. Four regression models (random forests) are trained to retrieve one parameter each, and they operate on manually-designed feature extractors obtained for an input HSI of size  $w \times h \times b$ , where  $w$  and  $h$  is its width and height, respectively, and  $b$  denotes the number of spectral bands (in this study,  $b = 150$  or  $b = 192$ ).

#### A.4 SHAP VISUALIZATIONS

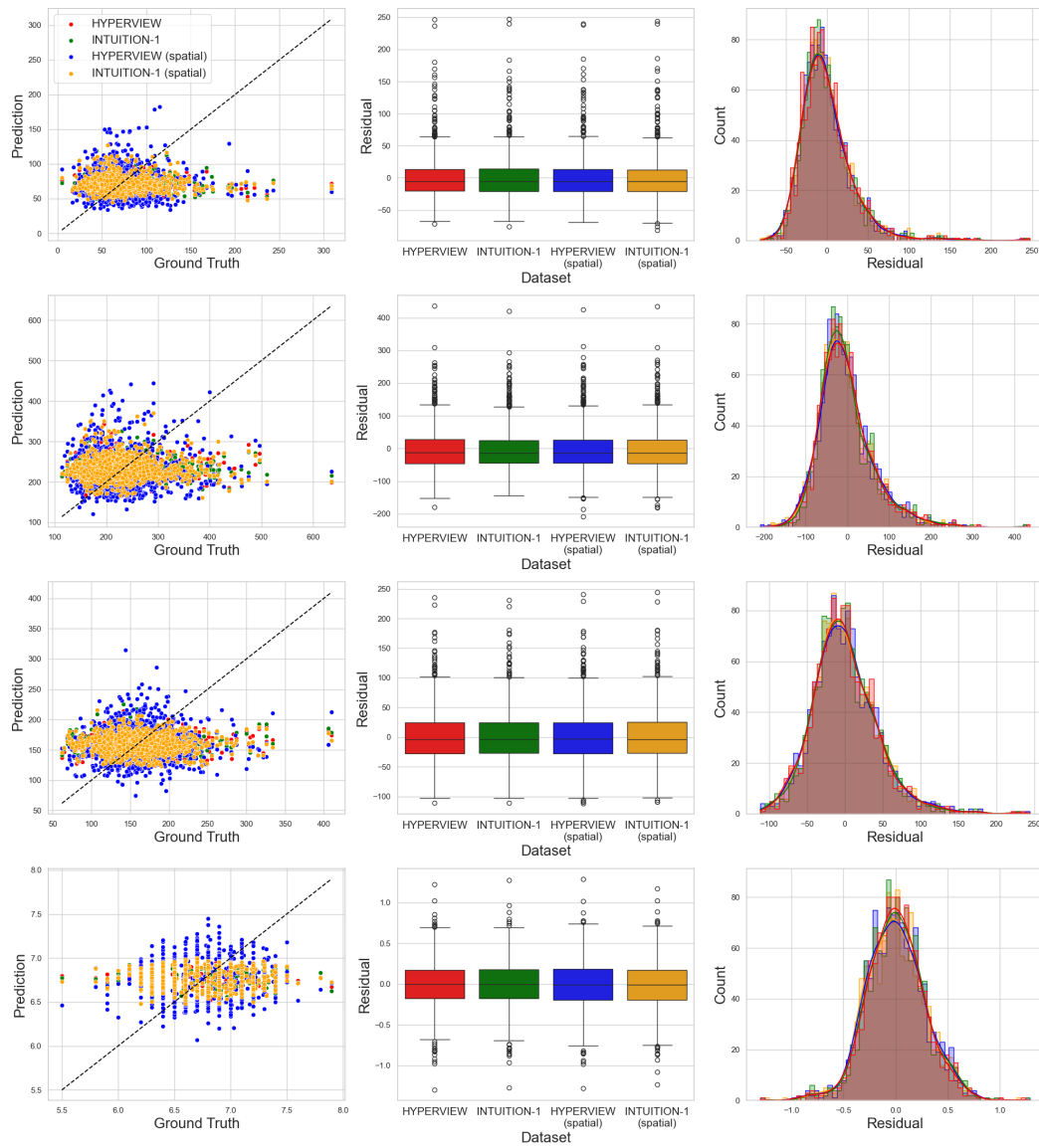


Figure 4: Residuals visualization for each soil parameter across different model versions. Each row represents a soil parameter, in the order of *phosphorus*, *potassium*, *magnesium*, and *pH*. Within each row: the left panel shows a scatter plot (ground truth on the  $x$ -axis, predicted values on the  $y$ -axis), the middle panel displays a boxplot of residuals, and the right panel presents a histogram of the residuals.

### A.5 SHAP AGGREGATION VISUALIZATIONS

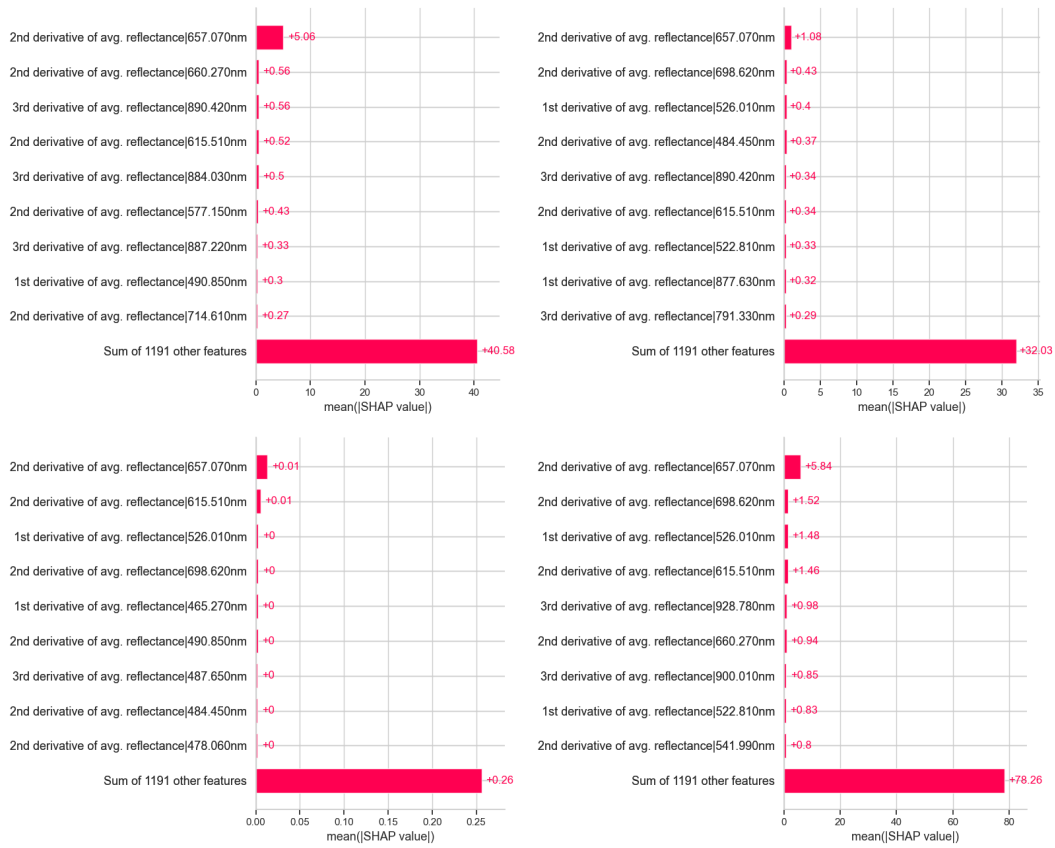


Figure 5: Visualization of the top 10 features identified by Shapley values, ordered by their contribution to model predictions in the EAGLEEYES model trained for the HYPERVIEW challenge. The panels show, clockwise from top left, *phosphorus*, *potassium*, *magnesium*, and *pH*.

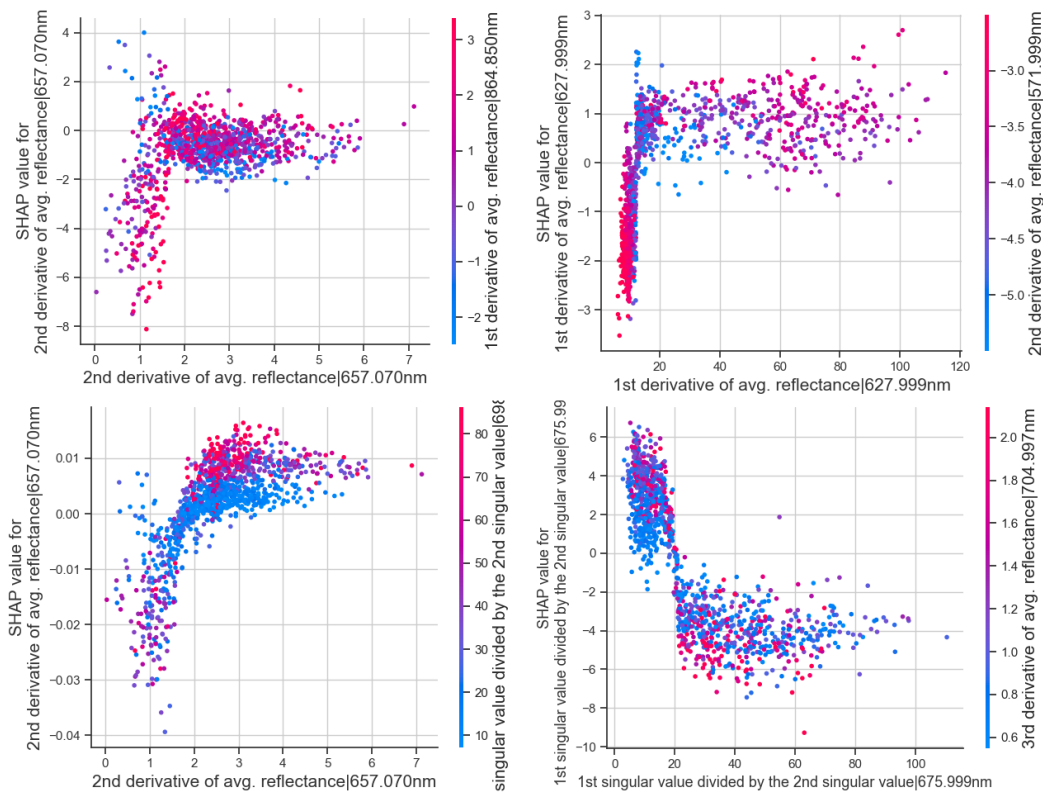


Figure 6: SHAP dependency plots showcasing the most influential feature in each iteration of the EAGLEYES model. The arrangement is as follows: HYPERVIEW (top left), INTUITION-1 (top right), HYPERVIEW (spatial) (bottom left), and INTUITION-1 (spatial) (bottom right). The  $x$ -axis represents the variation in feature value, while the  $y$ -axis depicts the corresponding change in Shapley values.

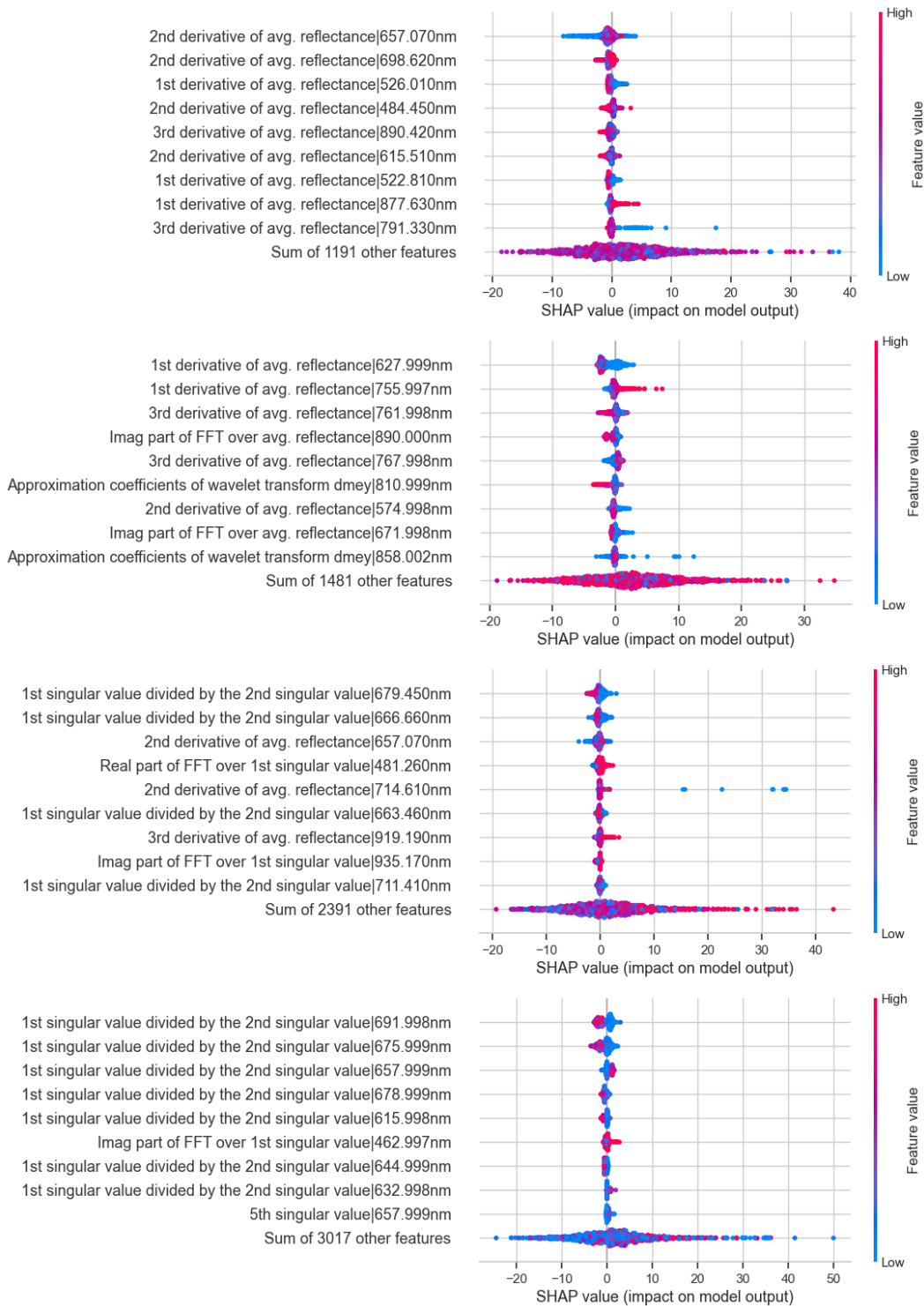


Figure 7: SHAP beeswarm plots for each iteration of the EAGLEEYES model, starting from the top HYPERVIEW, INTUITION-1, HYPERVIEW (spatial), and INTUITION-1 (spatial), displaying how feature values contribute to predictions. Features are displayed in descending order of influence.

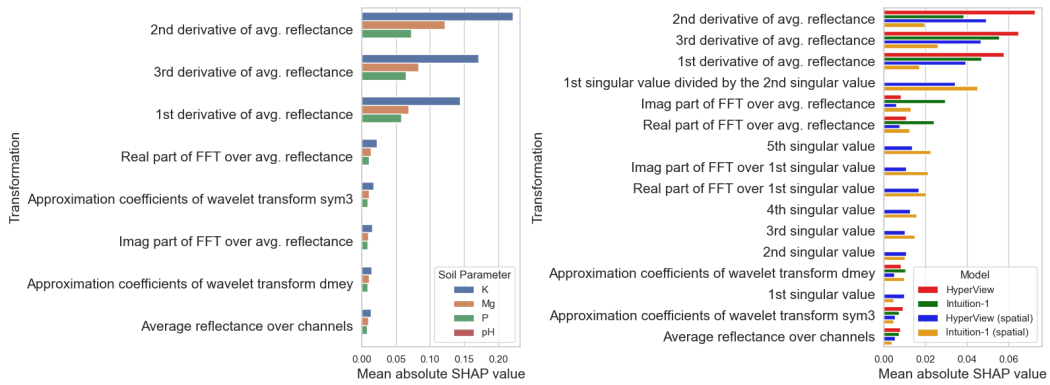


Figure 8: Shapley values for transformation aggregation in the EAGLEEYES models, showing feature importance across soil parameters (left) and model iterations (right).

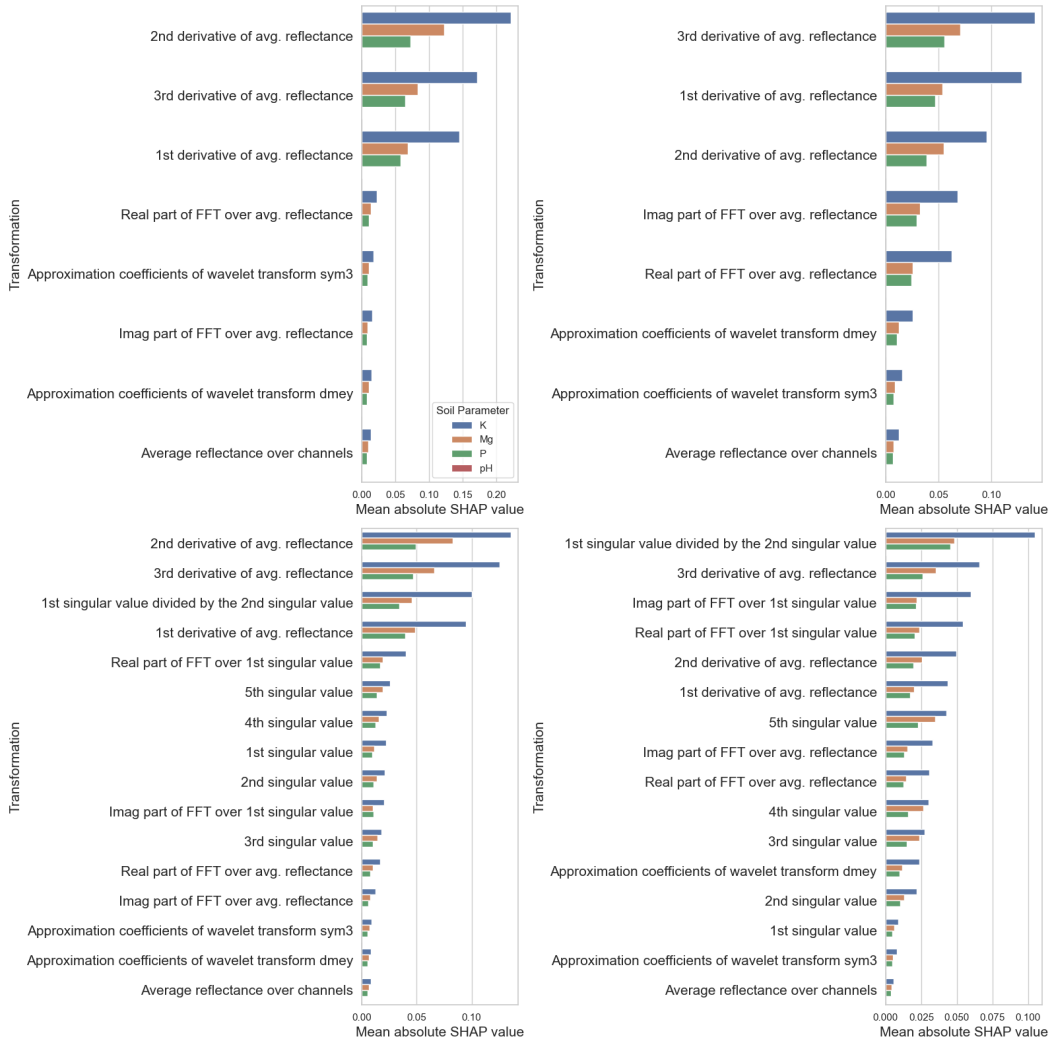


Figure 9: Shapley values for transformation aggregation across each iteration of the EAGLEEYES model, presented as follows: HYPERVIEW (top left), INTUITION-1 (top right), HYPERVIEW (spatial) (bottom left), and INTUITION-1 (spatial) (bottom right), illustrating feature importance across various soil parameters.