

Dear Editor,

We have now responded to all the reviewer and community comments, as summarised in depth below. The manuscript has been updated in several places in order to address their comments.

Kind regards,

Sam Thiele

Reviewer 1

I congratulate the authors on a well-executed and well-documented contribution to the science of predicting useful rock properties from proxy data.

Aside from some very minor grammatical corrections, highlighted in the attached pdf, I have a couple suggestions related to models:

Section 4.2 (Rock property prediction): earlier in the manuscript reference is made to 332 samples. I assume this constitutes the test/train dataset for the exercises described in this section. That should be specified, for clarity, and mention made of the percentage held back for testing or validation. 332 well curated samples is (relative to other geoscience regression problems for prediction of rock properties at least) a reasonable starting point. In the big scheme of things a productionizable set of models for predicting these characteristics would (as always) benefit from more training data. Please discuss this in-text here or under the Discussion heading.

In a similar manner, can you please specify if the datapoints shown in Figure 6 are the held back test data or the totality of the 332 samples after passing through your models. If the latter is true, I would strongly suggest recreating these figures with ONLY the holdback/test subset datapoints plotted, as a more realistic representation of how your models might behave in the wild.

With those suggestions incorporated I would be happy to recommend this for publication. Very nice work and clear explanations of complex subject matter.

We thank the reviewer for their feedback and are happy that they found our work interesting. We have integrated the grammatical corrections suggested in the pdf, and respond to the comment regarding the test split below.

Test/train split

As is relatively standard practice, we employed a 5-fold cross validation strategy in which five folds (each containing 20% of the data) were defined, and five (independent) models trained that each exclude a different fold as test data. This has been clarified in the caption for Fig. 6 (which shows only test-fold predictions), and by adding the following sentence to Section 3.4:

“Five models of each type were trained, each setting aside a single fold (20% of the data) as a test set. Each trained model was then used to predict its unseen test-set, and the results compiled for a robust assessment of model accuracy.”

We agree with the reviewer that our dataset of 332 samples is a good place to start exploring the relationship between hyperspectral data and mechanical properties, but that more would likely improve the accuracy (and be needed before applying this approach in e.g., an industrial setting). The final paragraph of Section 5.4 has been modified to better convey this:

“Finally, we caution that further development and the acquisition of a larger, more diverse training database is undoubtedly needed before this approach can be confidently applied to industrial

applications, especially for outcrop mapping. The lower-quality of hyperspectral data acquired outside of laboratory conditions and the variety of weathering processes that can influence outcrop surfaces, require approaches that are robust and carefully validated. However the required sensors and acquisition techniques already exist, suggesting cm-scale mapping of outcrop physical and mechanical properties is achievable, with appropriate site-specific calibration and validation.”

Reviewer 2

This study investigates the use of hyperspectral imaging to predict physical and mechanical properties of volcanic rocks (density, porosity, uniaxial compressive strength, and Young's modulus). By analyzing reflectance data across VNIR, SWIR, MWIR, and LWIR ranges, machine learning models, particularly multilayer perceptron (MLP), achieved high accuracy, explaining up to 80% of density and porosity variance and 65–70% of UCS and E variance. The research highlights the role of hydrothermal alteration, identifying spectral indicators for minerals like kaolinite and sulfates, and their impact on rock properties. The study also explores light-matter interactions, emphasizing surface and volume scattering effects. The findings demonstrate the scalability of hyperspectral imaging for remote sensing and outcrop mapping.

The quality of the manuscript is very high, and it makes a significant contribution to the field of geo-engineering characterization of volcanic rocks using hyperspectral remote sensing. The paper is certainly worthy of publication. I suggest that the authors elaborate a bit more on what is shown in Figures 7 and 8, to help readers better interpret and understand the results. It is recommended to clarify in the main text what the different colors and symbols in these figures represent, as they are not entirely self-explanatory.

We thank the reviewer for their assessment of our work, and have elaborated on Fig. 7 and 8 as suggested. Specifically, we have added the following sentences to the main text (Section 4.3) to better explain the SHAP results:

“Shapley values calculated for our ensemble predictions were aggregated to explore the contribution of each spectral range. This result exploits the additive nature of Shapley values: values derived for bands in the VNIR, SWIR, MWIR and LWIR ranges (respectively) can be summed to quantify the aggregate effect of each spectral range on each model prediction (Fig. 7). The results suggest the VNIR-SWIR range contributes most to predictions of density, UCS, and E that are below the expected (average) prediction, while the LWIR range makes a substantial contribution for above-average predictions. The opposite can be seen for porosity, where VNIR-SWIR bands mostly drive above average predictions. This pattern suggests the models learn to associate SWIR-active alteration minerals with reduced UCS, E, and density (and increased porosity).

The non-aggregated (per-band) Shapley values can also constrain the specific spectral features that, in combination, contribute to increase or decrease each prediction relative to the mean. These values are shown in Fig. 8, though only for models trained on the basaltic (Fig. 8a) and andesitic (Fig. 8c) subsets separately (to reduce the influence of lithological effects). The results are difficult to interpret specifically because the predictions result from a complex balance between positive contributions from some bands (red) and negative contributions (blue) from others. Strongly negative Shapley values are often associated with 1800, 1900, and 2200 nm bands, which contain absorptions characteristic of hydrothermal alteration minerals (Table 1) for samples with low predicted E. Higher predictions also appear driven by these same bands, possibly due to an absence of absorption features in these wavelengths for these samples. In the MWIR, features at ~3400 and between 4200 and 4900 nm appear important, with several “doublets” (spectrally adjacent high and low Shapley values) indicating a sensitivity to absorption shape (asymmetry) or position. The first of these bands (3400 nm) is likely related to ν₂HOH absorptions (though this absorption will have been heavily distorted by the hull correction applied during pre-processing). The latter bands (4200–4900) are interpreted to relate to 2νSi-O absorptions from silicate minerals or 2νS-O absorptions from sulphates (Laukamp et al., 2021). The last of these (4900) may also have been shifted by the hull correction.

The Shapley values are easier to interpret after averaging their absolute value across all samples, to broadly highlight important spectral ranges. As mentioned also above, these ranges (Fig. 8b and Fig. 8d) match several expected mineralogical absorptions but, interestingly, also suggest that the model tends to focus on absorption “shoulders” rather than their centres, which we speculate could be due to a higher sensitivity of absorption shoulders to complex scattering effects.”

Community Comment 1

I read the abstract with interest, and briefly scanned through the manuscript from a geomechanical perspective. In section 5.2, it is stated there is a correlation between UCS strength and different spectral measurements, without a theoretical underpinning why that might be the case. Without such an underpinning it remains a correlation, which doesn't mean there is a causal relationship. Given that weakness, the abstract needs to be toned down in terms of certainty. Correlation is not causation. There is some reasoning why density would be related to spectral information, so it could equally be possible to outline reasons why there would be a causal relationship.

Thank you for the suggestion, and interest in our work. It is correct that we note in several places (including 5.2) that there is a correlation between UCS and hyperspectral response, but (1) certainly do not imply that this is a causal correlation (quite the opposite in fact), and (2) we discuss the theory behind why such a correlation might exist in quite some depth within Section 2 (theory).

We agree that correlation is not causation, but we do not see this as a problem because we are not trying to imply a causal relationship, but rather develop hyperspectral as a useful and relatively easy to collect proxy variable. This logic is succinctly summarised in the abstract: *“These results demonstrate that hyperspectral imaging can serve as a robust proxy for rock physical and mechanical properties, potentially offering an efficient, scalable method for characterising large areas of exposed volcanic rock.”*