# Enhancing Soil and Tailings Characterization: Integration of Photonics Technology with Cone Penetration Testing

**Iman Entezari**, ConeTec Group, Canada
**Dallas McGowan**, ConeTec Group, Canada
**Joseph Bindner**, ConeTec Group, Canada
**Joseph Glavina**, ConeTec Group, Canada
**Jason DeJong**, University of California, Davis, USA

## Abstract

The integration of photonics technology, particularly hyperspectral sensing and high-resolution imaging, with piezocone penetration tests (CPTu) offers a transformative approach to soil and tailings characterization. This paper presents preliminary work on the assessment of hyperspectral data combined with CPTu data for tailings characterization through analyzing a large dataset of paired hyperspectral, CPTu, and laboratory results using machine learning models. Additionally, it presents discussion on initial findings on particle size estimation employing a machine learning based approach to analyze high-resolution imaging acquired from reference samples and in-situ tests. The results demonstrate that incorporating hyperspectral and CPTu data significantly enhances tailings characterization, and the particle size determination using high-resolution images is promising.

## Introduction

In geotechnical and geoenvironmental site investigations, characterization of soil and soil-like properties is essential. This often requires a combination of in-situ testing and sampling methods. Traditionally, laboratory testing and piezocone penetration testing (CPTu) have been the common techniques for soil characterization. While laboratory testing offers a comprehensive analysis of various soil properties, from constitutive, to geotechnical, to chemical, it involves drilling, sampling, and subsequent transportation, which is costly, carbon-intensive (Purdy et al., 2021), and more hazardous than CPTu operations. In-situ collection of soil hyperspectral data and high-resolution images coupled with CPTu or Seismic CPTu (SCPTu) offers promise for reducing cost, time, and environmental impacts associated with sampling and laboratory testing.

Building upon the foundation of CPTu technology, advancements are being made in incorporating hyperspectral sensors and high-resolution cameras into modules that can be placed in-line with the CPTu. The hyperspectral module enables real-time and continuous spectral data collection across an extended wavelength spectrum during CPTu operations. The vision module, designed for RGB imaging, facilitates the acquisition of high-resolution photographs in real-time during the CPTu push.

The hyperspectral module provides valuable insights into the chemical composition and crystal structure of geomaterials by measuring reflected light across different wavelengths. This capability allows for the identification and quantification of minerals (Clark, 1999; Entezari et al., 2017; Debaene et al., 2023), estimation of water content (Fabre et al, 2015; Entezari et al., 2016), detection and quantification of contaminants (Schwartz et al., 2011; Shi et al., 2014), and assessment of chemical composition (Cozzolino and Moron, 2003; McBride, 2022). The vision module provides detailed texture and color information about the geomaterial, aiding in the estimation of particle size distribution and identification of soil types (Hryciw et al., 2009 and 2015; Ventola and Hryciw, 2023).

This paper presents preliminary work on the analysis and interpretation of hyperspectral data coupled with CPTu data for tailings characterization. This assessment is conducted through the analysis of a dataset of paired hyperspectral and CPTu data, alongside tailings properties from laboratory testing. In our previous studies, we have developed machine learning models for the estimation of tailings constituents using hyperspectral data (Entezari et al., 2022a) and using Gamma-CPTu (GCPTu) data (Entezari et al., 2022b). In this paper, we evaluate the potential of combining hyperspectral and CPTu data for improved tailings characterization. Additionally, we present initial pilot results on particle size estimation using high-resolution benchtop and in-situ images analyzed with a machine learning based algorithm.

## Background

### Hyperspectral module

The hyperspectral module (patent pending) is configured with spectrometers capable of collecting reflectance spectra in the spectral ranges of 400-850 nm and 950-1700 nm. Alternate configurations can be used to collect different spectral ranges. To collect the reflectance spectra during CPTu operations, the hyperspectral module contains a light source to illuminate the soil through a window. The reflected light is then transferred to a spectrometer(s) that transmits a digitized signal to the CPTu data acquisition system (DAS). Figure 1 illustrates sample data obtained from a wet and a dry tailings sample using a benchtop setup of the hyperspectral module.

By measuring reflected light across different wavelengths, spectral sensors provide information about chemical composition and crystal structure (Clark, 1999). This versatile method enables both qualitative and quantitative assessments of various materials and substances.

Reflectance spectroscopy facilitates quantitative assessments of soil properties by correlating spectral signatures with specific parameters of interest. Through empirical relationships established between spectral features and target properties, quantitative estimations with high accuracy can be derived. By calibrating spectral data against reference measurements obtained from laboratory analyses, robust quantitative models can be constructed for diverse soil properties, including mineral content, organic carbon content, and contamination levels.

In qualitative analysis, reflectance spectroscopy can help identify and characterize different components present in the soil. By analyzing the spectral signature (the unique pattern of light absorption and reflection across various wavelengths) the presence of minerals, organic matter, and contaminants can be identified (Clark, 1999; Schwartz et al., 2011). Each component exhibits distinct spectral features, allowing for their identification without the need for physical extraction or laboratory testing.

The integration of hyperspectral sensors into CPTu modules allows for real-time, continuous hyperspectral reflectance data collection during CPTu or SCPTu operations. The synergistic integration of hyperspectral data with CPTu data (or SCPTu), correlated and analyzed by machine learning algorithms, can produce new methods to estimate soil properties such as plasticity, fines content, water content, degree of saturation, mineralogy, and contaminants.

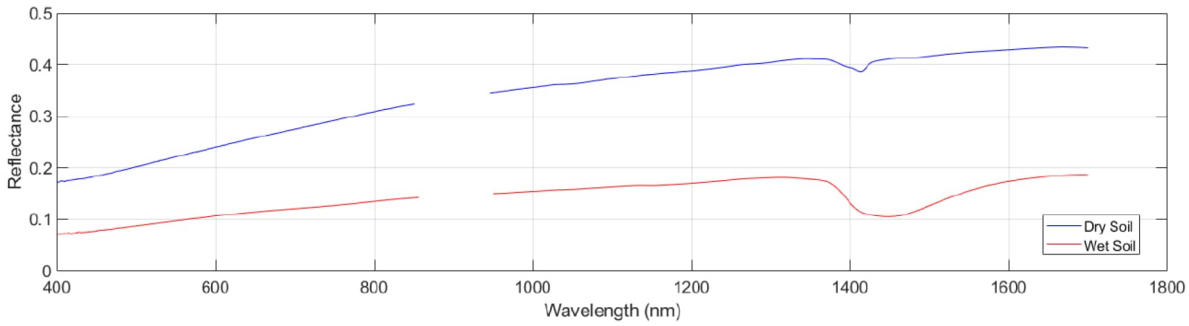


**Figure 1: Example hyperspectral data from hyperspectral module**

## Vision module

The vision module is equipped with an 18 MP high-resolution image sensor which captures photographs through a 16 mm diameter window. With the present configuration, each pixel of the image corresponds to an area of approximately 4µm × 4µm. Figures 2a and 2b show example images acquired from two soil samples with particle diameters ranging from150-250 µm and 105-150 µm, respectively.

Vision module offers numerous advantages for geotechnical and geo-environmental site characterization. The ability to perform visual inspections using in-situ images reduces the need for visual assessment by physical soil sampling and complements soil behaviour type (SBT) derived soil stratigraphy. Additionally, it facilitates a thorough quality check of CPTu data by pairing images with the data, allowing for the identification of anomalies through supplementary visual evidence. This technology can also enable

the estimation of particle size distribution through textural analyses or machine learning modelling (Buscombe, 2019; Ventola and Hryciw, 2023). Furthermore, the detailed insights gained from analyzing particle size and colors enhance soil classification and characterization, making vision CPTu a valuable tool in site assessment and analysis.

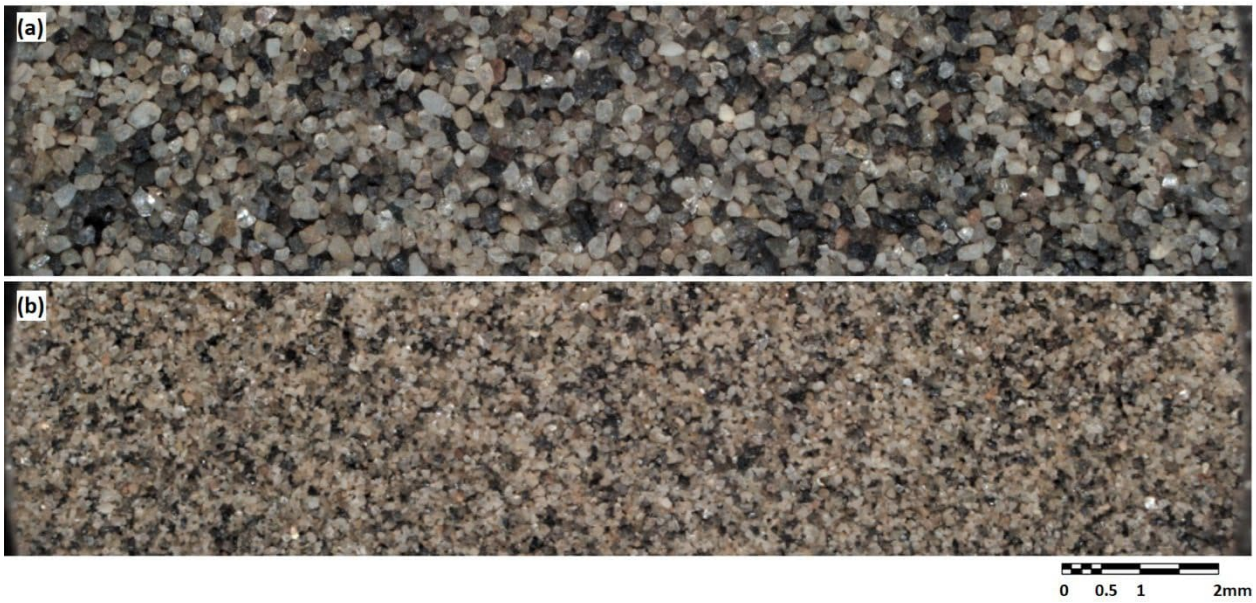


**Figure 2: Example vision module data from samples with particle size of (a) 150-250 µm and (b) 105-150 µm**

## Data and methods

### Hyperspectral GCPTu dataset

To explore the potential synergy between hyperspectral and CPTu data, a dataset combining hyperspectral and GCPTu data from the oil sands region in Canada was compiled. ConeTec's oil sands tailings dataset (Entezari et al., 2022a), originally collected in the 350-2500 nm range from ex-situ samples, was limited to the spectral ranges of 400-850 nm and 950-1700 nm, corresponding to the spectral ranges of the hyperspectral module.

To pair hyperspectral data with GCPTu data, our geospatial database was queried to identify GCPT soundings in proximity (within 10 m radially, 3 months temporally) of the hyperspectral data samples. The GCPT parameters, including gamma counts, undrained shear strength, and the slope and linearity of porewater pressure (Entezari et al., 2022b), were averaged over the sample length and paired with hyperspectral data. This process resulted in a dataset comprised of hyperspectral data, GCPTu data, and corresponding laboratory results. The final dataset included 4723 data pairs, providing a large dataset for analyzing the combined utility of hyperspectral and CPTu data in tailings characterization. The dataset was split into training and test sets, through random selection of ~10% of the data points as the test set. Four

models were trained using four different sets of input features including: (1) hyperspectral and GCPTu parameters, (2) hyperspectral and CPTu parameters (no gamma), (3) only hyperspectral parameters, and (4) only GCPTu parameters. Consistent with previous work (Entezari et al. 2022a and 2022b), the models of this study were trained using the training set and an ensemble of neural networks using bootstrap aggregation (Bagging) technique (Sollich and Krogh 1996, Breiman 1996). The models were trained to predict the contents of bitumen, solids, water, total fines, fines/(fines+water) (FFW) as well as methylene blue index (MBI). The performance of the models was evaluated using the cumulative distribution function (CDF) of errors on the test set (Error = Lab -predicted). The bias of the prediction was determined at the 50$^{th}$ percentile in the CDF. Assuming normal distribution of errors, the CDF values at 15.9% and 84.1% correspond to ±1 standard deviation and the average of these CDF values was adopted as the comprehensive error measure for the model.

### High resolution images

To assess the use of vision module images to determine particle size distribution in coarse grained soils, the Segment Every Grain machine learning procedure (Sylvester, 2023) was applied to two images: one from a benchtop study, and one from in-situ imaging. The Segment Every Grain algorithm detects particles in images and provides estimates of particle diameter by identifying boundaries and determining the major and minor axis length of particles. The algorithm first identifies particles, then allows the user to delete improperly detected particles and add new particles. This combined machine-user approach allows for quick refinement of modeled results.

Imagery was first captured on the reference sample with particles ranging from 105 -150 μm (Figure 2b). The algorithm was then applied to the image and results were refined by the user. A field demonstration was also conducted using the vision CPTu to collect imagery and CPTu data up to a depth of 12 meters in oil sands tailings. During the CPTu test, the probe was stopped at 10 cm increments to capture images. Data collected from in-situ tests were compiled and a sample image of a predominantly sandy material was selected to assess the performance of the Segment Every Grain algorithm for the particle size estimation of the in-situ images.

## Results

### Tailings characterization using hyperspectral and CPTu data

The results of the trained models developed using different sets of input features are presented in Figures 3-6. These figures include scatter plots that depict the relationship between lab measured constituents and the predicted results for the test set. The model architecture and training parameters were consistent across all four models to facilitate a comparative analysis. As shown, models incorporating both hyperspectral and

GCPTu/CPTu parameters as input features outperformed those trained solely with hyperspectral parameters or solely with GCPTu parameters across all tailings characterizations. The summary of the error analysis is provided in Table 1.

The application of the developed models on in-situ data faces two challenges. First, collecting hyperspectral data in oil sands fluid tailings is impractical due to bitumen smearing and the lack of a self-cleaning mechanism, limiting the application of the model to soil-like materials in the oil sands where data collection is feasible. However, this limitation does not apply to non-hydrocarbon deposits. An error analysis was therefore conducted exclusively on soil-like material where in-situ hyperspectral data collection would be feasible (Table 1). The soil-like material was identified based on a pore pressure linearity of less than 0.8.

Second, variations between the spectrometers used during model development and those in the hyperspectral module may require calibration transfer to ensure accurate data interpretation. Future efforts will evaluate the need for calibration transfer methods to align data from different spectrometers, thereby enhancing the model's reliability and applicability to in-situ data.

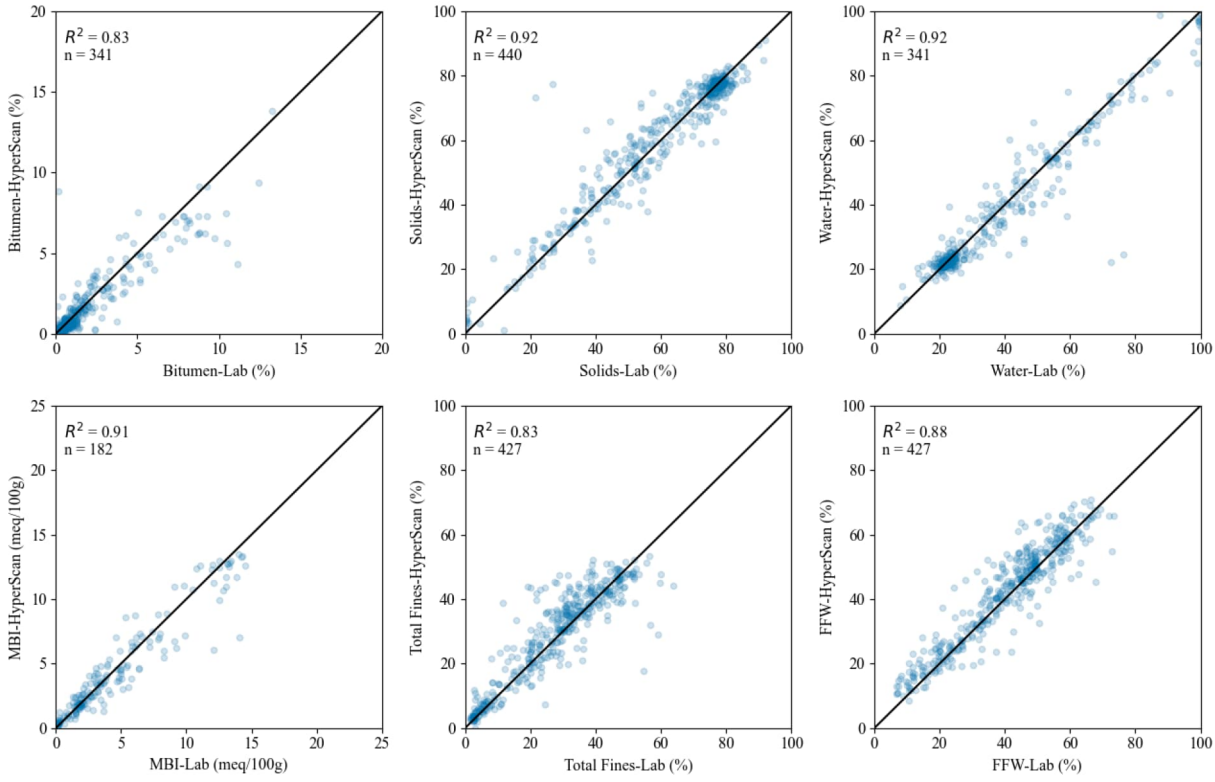


**Figure 3: Performance assessment on the test set using the model trained with both hyperspectral and GCPTu parameters as input features**

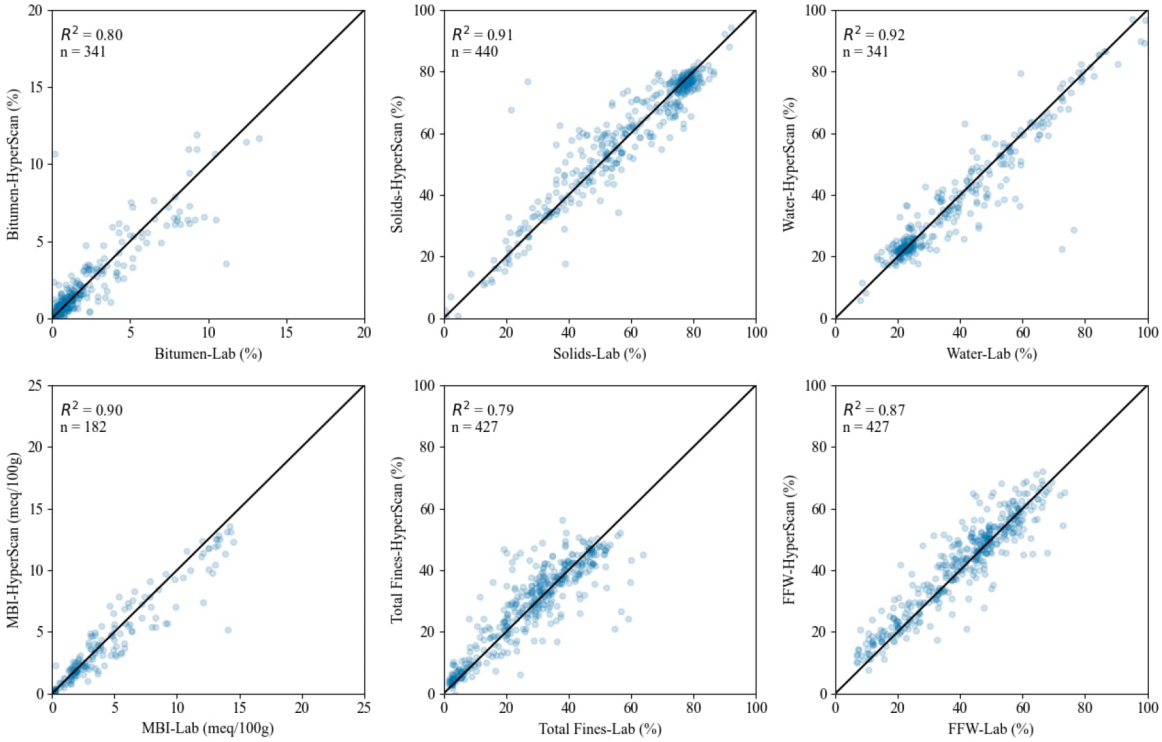


**Figure 4: Performance assessment on the test set using the model trained with both hyperspectral and CPTu parameters as input features**

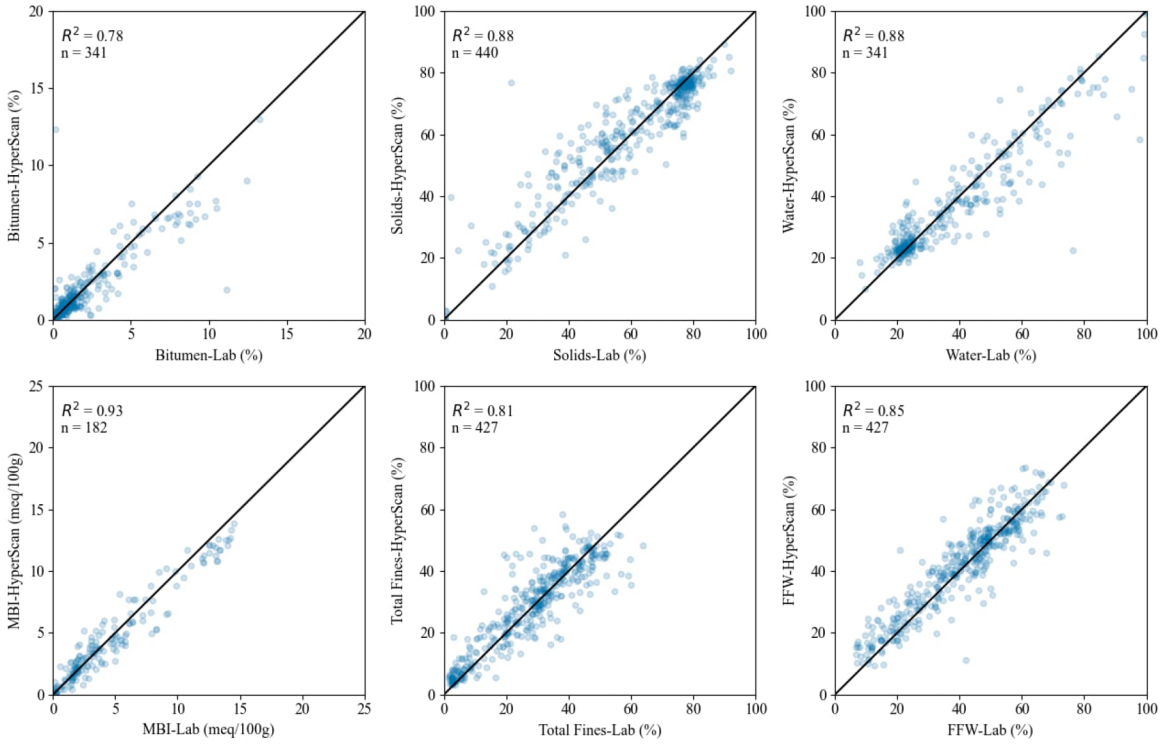


**Figure 5: Performance assessment on the test set using the model trained with only hyperspectral data as input features**

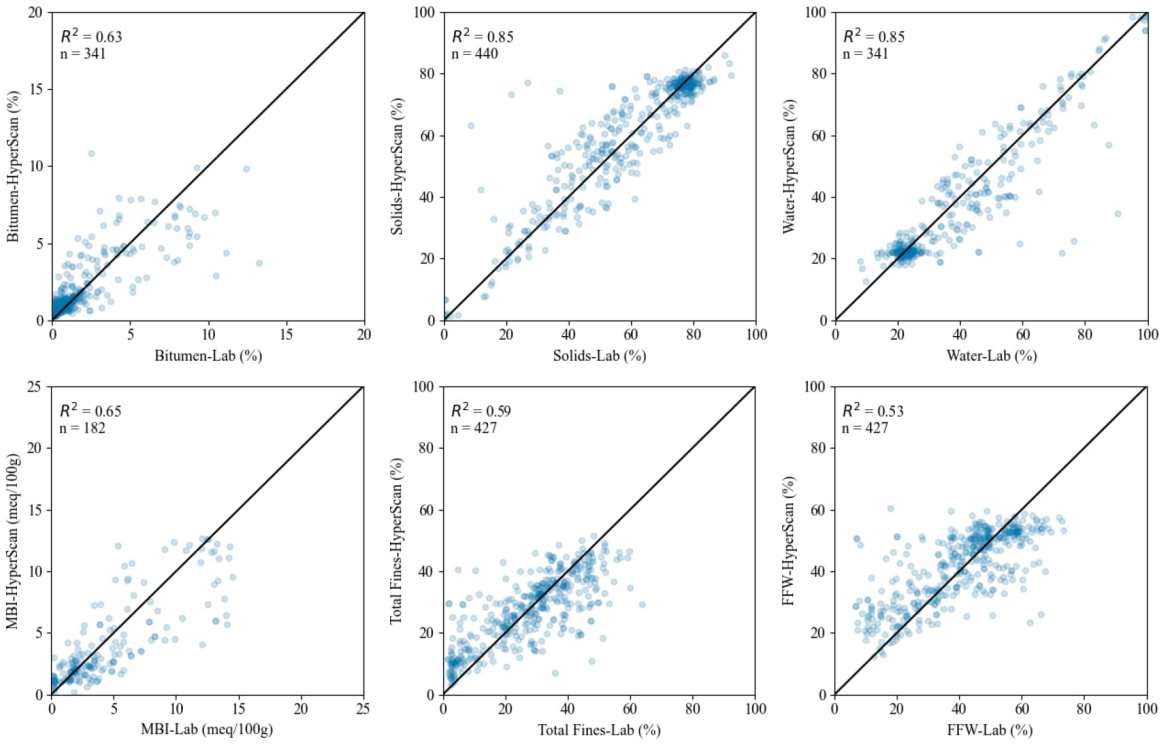


**Figure 6: Performance assessment on the test set using the model trained with only GCPTu data as input features**

**Table 1: Summary of error analysis (bias ± error)**

| | Hyperspectral+GCPTu | Hyperspectral+CPTu | Hyperspectral Only | GCPTu Only |
|---|---|---|---|---|
| | **Entire Test Set** | | | |
| Bitumen (wt%) | 0.05 ± 0.49 | -0.08 ± 0.49 | -0.03 ± 0.5 | -0.25 ± 0.69 |
| Solids (wt%) | -0.53 ± 4.18 | 0.15 ± 4.43 | 0.27 ± 5.59 | 0.01 ± 5.88 |
| Water (wt%) | 0.45 ± 3.2 | 0.1 ± 3.7 | -0.26 ± 5.15 | 0.73 ± 5.8 |
| MBI (meq/100g) | 0.01 ± 0.88 | 0.11 ± 0.92 | 0.09 ± 1.02 | 0.2 ± 1.91 |
| Total 44 µm Fines (wt%) | -0.99 ± 4.51 | -0.98 ± 4.8 | -1.0 ± 4.84 | -0.98 ± 8.17 |
| FFW (wt%) | -1.67 ± 4.29 | -1.19 ± 4.4 | -1.41 ± 5.04 | -1.53 ± 8.86 |
| | **Soil-Like Tailings fraction of Test Set (linearity of pore pressure < 0.8)** | | | |
| Bitumen (wt%) | 0.02 ± 0.34 | -0.07 ± 0.33 | -0.07 ± 0.31 | -0.26 ± 0.47 |
| Solids (wt%) | -0.09 ± 2.85 | 0.43 ± 2.98 | 1.37 ± 3.21 | 0.13 ± 4.19 |
| Water (wt%) | 0.04 ± 2.71 | -0.18 ± 2.82 | -1.27 ± 3.2 | 0.69 ± 4.3 |
| MBI (meq/100g) | -0.06 ± 0.43 | 0.04 ± 0.61 | -0.1 ± 0.66 | -0.09 ± 1.33 |
| Total 44 µm Fines (wt%) | -0.73 ± 4.79 | -1.07 ± 5.75 | -1.37 ± 5.84 | -1.78 ± 11.33 |
| FFW (wt%) | -1.54 ± 4.94 | -1.55 ± 4.93 | -1.19 ± 6.29 | -0.54 ± 13.11 |

### PSD analysis using images

Figure 7a presents a benchtop image of a reference sample captured by the vision module, with the highlighted area enlarged in Figure 7b. The particle detection results are shown in Figure 7c, where the algorithm successfully detected most particles. The PSD results from machine learning are displayed in Figure 7d. Most particles fall within the 105-150 µm range (#100-140 sieve), though sizes below 105 µm and above 150 µm are also observed (Figure 7d). A close examination of Figure 7c, compared with the grains in Figure 7b, indicates that some adjacent grains were detected as larger agglomerated particles, leading to sizes over 150 µm. Additionally, overlapping particles sometimes resulted in only the uncovered part of a particle being detected, accounting for sizes under 105 µm.

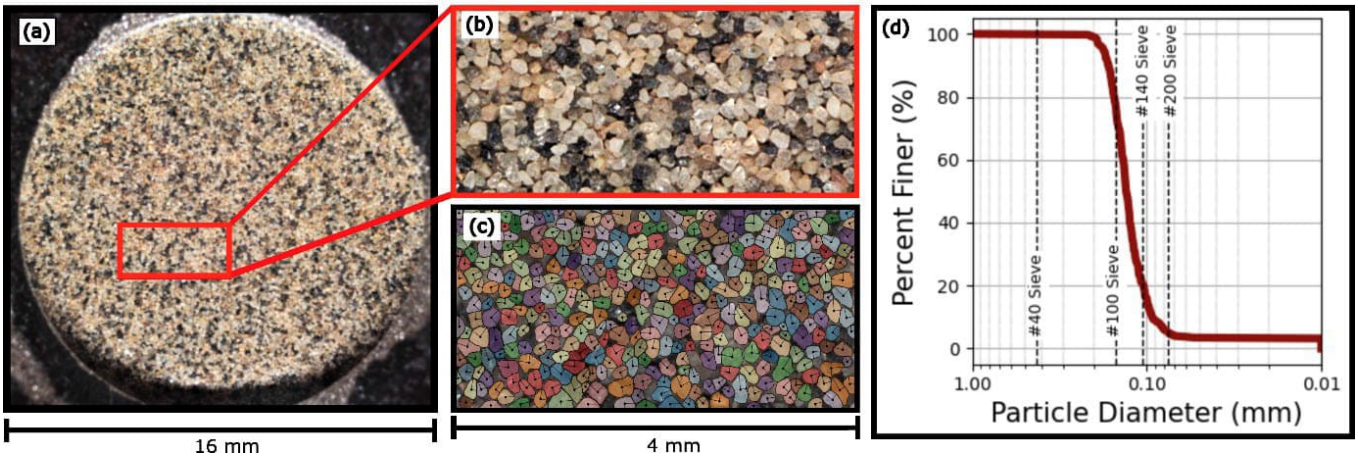


**Figure 7: Results from benchtop imaging including (a) the image of the reference sample, (b) the cropped section used in analysis, (c) results from machine learning segmentation, and (d) the particle size distribution curve for detected particles**

Figure 8 presents the results from in-situ field testing using the vision module, showcasing cone tip resistance (qt), sleeve friction (fs), stacked incremental images compiled from cropped thin sections from the imagery, and enlarged image sections at selected depths. As shown in Figure 8, a transition in qt is observed near a depth of 7 m, which corresponds to a transition from predominantly sandy material to fine grained mixtures. When cross referencing this information with in-situ images, a color change in the tailings profile is observed. Additionally, a visual change in moisture condition above and below the water table at 0.6 m depth is observed based on the enlarged images.

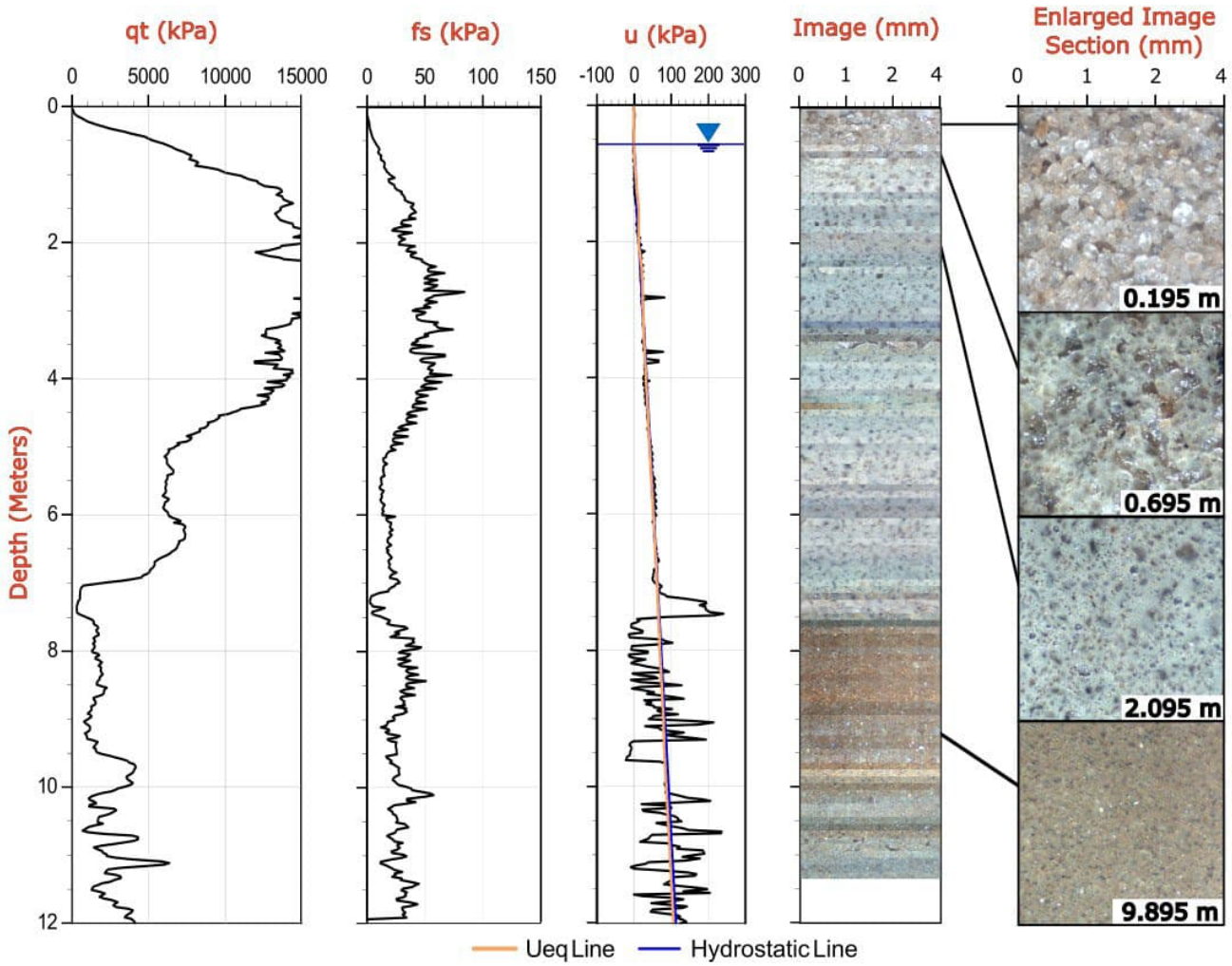


**Figure 8: CPTu data and associated images including the cone tip resistance (qt), sleeve friction (fs), dynamic pore pressure (u), cropped image at each imaging depth, and enlarged image sections at selected depths**

Figure 9 illustrates the results of the Segment Every Grain algorithm applied to an in-situ image from sand-dominated tailings. Visual interpretation suggests that the algorithm successfully identified the majority of particles in the image. However, due to the similar color of the particles, its performance is limited. While laboratory data are required to thoroughly evaluate the accuracy of the results, the preliminary PSD analysis indicates that most particles are larger than 75 μm, which is typical for such tailings. This initial assessment shows promise for predicting particle size distribution from in-situ images.

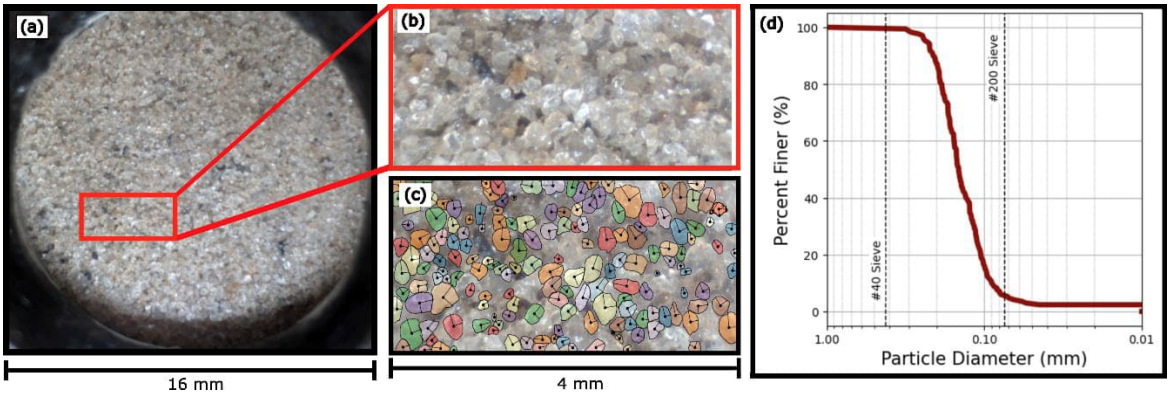


**Figure 9: Results from in-situ imaging including (a) the image of the sand-dominated tailings, (b) the cropped section used in analysis, (c) results from machine learning segmentation, and (d) the particle size distribution curve for detected particles**

## Conclusion

This paper introduced initial research on the integration of advanced photonics technologies, including hyperspectral sensing and high-resolution imaging, into CPTu modules. Results show that the synergistic fusion of hyperspectral data with CPTu data, empowered by machine learning algorithms, improves the accuracy and reliability of predicting tailings properties. Analysis of a dataset containing hyperspectral, CPTu, and laboratory data demonstrates that modeling procedures that incorporate both hyperspectral and GCPTu/CPTu data outperform models that only use hyperspectral or CPTu data. Additionally, the use of the high-resolution images and machine learning-based analysis shows promise for the estimation of particle size distribution for coarse grained materials. Although the application of optical technologies and machine learning for soil and tailings characterization is still a developing field, this study, among others, highlights the potential of these technologies to enhance the speed, cost-efficiency, and resolution of site characterization.

## References

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24(2): 123–140.

Buscombe, D. 2019. SediNet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry. *Earth Surface Processes and Landforms* 45(3): 638–651.

Clark, R.N. 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of Remote Sensing* 3: 3–58.

Cozzolino, D. and A. Moron. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* 140(1): 65–71.

Debaene, G., P. Bartmiński and M. Siłuch. 2023. In situ VIS-NIR spectroscopy for a basic and rapid soil investigation. *Sensors* 23: 5495.

Entezari, I., D. McGowan, J. Glavina and J. Sharp. 2022b. Tailings behaviour type to predict oil sands tailings constituents from the gamma cone penetration test. *Proceedings of the International Oil Sands Tailings Conference (IOSTC)*, Edmonton, AB: 6–15.

Entezari, I., D. McGowan and J. Glavina. 2022a. Hyperspectral imaging technology for oil sand tailings characterization: practical aspects. *Proceedings of Tailings and Mine Waste 2022*, Denver, CO, USA.

Entezari, I., B. Rivard, M. Geramian and M.G. Lipsett. 2017. Predicting the abundance of clays and quartz in oil sands using hyperspectral measurements. *Int. J. Appl. Earth Obs. Geoinf.* 59:1–8.

Entezari, I., B. Rivard, M.G. Lipsett and G.W. Wilson. 2016. Prediction of water content and normalized evaporation from oil sands soft tailings surface using hyperspectral observations. *Canadian Geotech. J.* 53(10): 1742–1750.

Fabre, S., X. Briottet and A. Lesaignoux. 2015. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4–2.5 µm domain. *Sensors* 15(2): 3262–3281.

Hryciw, R.D., H.-S. Ohm and J. Zhou. 2015. Theoretical basis for optical granulometry by wavelet transformation. *J. Comput. Civ. Eng.* 29(3): 04014050.

Hryciw, R.D., Y. Jung, E. Susila and A. Ibrahim. 2009. Thin soil layer detection by VisCPT and FEM simulations. *Proceedings of the 17th International Conference on Soil Mechanics and Geotechnical Engineering (ICSMGE)*, IOS, Amsterdam, Netherlands: 1052–1055.

McBride, M.B. 2022. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. *Eur. J. Soil Sci.* 73: e13192.

Purdy, C.M., A.J. Raymond, J.T. DeJong, A. Kendall, C. Krage and J. Sharp. 2021. Life cycle sustainability assessment of geotechnical site investigation. *Canadian Geotech. J.* 59 (6).

Schwartz G., G. Eshel and E. Ben-Dor. 2011. Reflectance spectroscopy as a tool for monitoring contaminated soils. In: Pascucci S., editor. *Soil Contamination.* New York: InTech: 67–90.

Shi, T., Y. Chen, Y. Liu and G. Wu. 2014. Visible and near-infrared reflectance spectroscopy – An alternative for monitoring soil contamination by heavy metals. *J. Hazard. Mater.* 265: 166–176.

Sollich, P. and A. Krogh. 1996. Learning with ensembles: how over-fitting can be useful. In *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge, MA: 190–196.

Sylvester, Z. 2023. Segmenteverygrain. Accessed at https://github.com/zsylvester/segmenteverygrain

Ventola, A. and R.D. Hryciw. 2023. An autoadaptive Haar wavelet transform method for particle size analysis of sands. *Acta Geotech.* 18: 5341–5358.