Efficient Earth Observation System for Acid Mine Drainage

Veronika Kopačková-Strnadová[1], Martin Kýhos[1], Jan Jelének[1]

[1]*Czech Geological Survey, Prague 11821, Czech Republic, veronika.strnadova@geology.cz, https://orcid.org/0000-0001-9317-5990*

Abstract

Acid Mine Drainage (AMD), a significant water pollutant from sulfide mineral oxidation in mining areas, generates sulfuric acid and harmful substances. Effective AMD management involves careful material handling, treatment, and water management. Imaging spectroscopy provides a practical alternative to traditional chemical analysis, identifying pyrite oxidation "hot spots" and sulfate mineral formation. This research develops robust AMD monitoring systems using machine learning techniques on optical multi- and hyperspectral data. We use hyperspectral (PIKA L camera, UAV) and multispectral (Sentinel-2, orbital) datasets at the Lítov post-mine dump (Sokolov lignite basin, Czech Republic). Validated against XRD mineralogy and Google Earth-identified hotspots, Radial Basis Function Support Vector Machine (RBF SVM) outperforms other ML methods in identifying AMD hot spots and class separation. RBF SVM effectively detects AMD discharge and distinguishes mineral mixtures (oxy-hydroxides and oxides) using both Pika L and Sentinel-2 data.

Keywords: Acid Mine Drainage, multispectral data, hyperspectral data, Machine Learning, classification

Introduction

Acid Mine Drainage (AMD) is a significant environmental issue that arises when water flows over or through sulfur-bearing rocks, such as coal or metal ores, which have been exposed to air and water during mining operations. This exposure leads to the oxidation of sulfide minerals, producing sulfuric acid. The generated acid reacts with surrounding rocks, releasing various toxic metals and metalloids, including arsenic, lead, and cadmium, into the water. Consequently, the discharged water becomes highly acidic and contaminated, posing severe ecological risks. Acidic water can devastate aquatic habitats, harm fish and other aquatic organisms, and significantly degrade overall water quality. It can also contaminate drinking water sources, leading to adverse effects on human health, including neurological and developmental issues. Addressing AMD requires implementing appropriate measures during mining operations, such as effective handling and treatment of mined materials and the establishment of robust water management systems. Additionally, post-mining remediation techniques, such as neutralizing the acid and removing harmful substances, are employed to minimize environmental impact and restore affected ecosystems.

Recent advancements in remote sensing technologies have revolutionized the monitoring of AMD. Imaging spectroscopy serves as an efficient alternative to traditional chemical analyses for mine characterization and assessing potential AMD discharge, as well as for acid sulfate soil mapping. Spectroscopic approach focuses on identifying minerals that indicate subaerial oxidation of pyrite (e.g., jarosite), often referred to as "hot spots," and the subsequent formation of the other oxidation products (such as oxy-hydroxides and oxides) (Swayze *et al.* 2000).

To date, effective methods have been demonstrated to detect potential AMD hotspots using hyperspectral systems deployed on Unmanned Aerial Vehicles (UAVs, Flores *et al.* 2021) and aerial platforms (Kopačková

2014). Moreover, the availability of current satellite systems that provide free multispectral data, such as Sentinel-2, and hyperspectral imaging data from platforms like EnMap and PRISMA, offers novel opportunities for monitoring AMD from space. Furthermore, a feasibility study by Chalkey *et al.* (2023) highlighted the potential of multi-scale AMD monitoring using various remote sensing platforms, including UAV-based systems, PlanetScope, and Sentinel-2. Their research emphasized the importance of integrating multiple datasets for comprehensive monitoring, which can enhance the accuracy and efficiency of AMD detection.

In this study, we aim to advance the development of innovative and highly efficient AMD monitoring systems by exploring the application of Machine Learning (ML) techniques to imaging spectral data acquired by different platforms, including UAV-based hyperspectral data (PIKA L) and multi-spectral imagery from Sentinel-2 satellite. By leveraging these technologies, we aim at improving the detection and management of AMD, ultimately contributing to more sustainable mining practices and better protection of the environment.

Test site

The study was conducted at the Lítov post-mine dump, located in the western part of the Sokolov lignite basin in the Czech Republic. This site is notable for its inclusion in the Czech-Bavarian "Geopark" program, which highlights the unique characteristics of man-made landscapes. The Lítov dump is particularly remarkable due to its highly acidic substrates (Kopačková 2014, Kopačková and Hladíková 2014), sparse vegetation, and the presence of a unique semi-desert environment that supports exceptional flora and fauna.

Remote sensing data

In September 2023, **UAV-based hyper-spectral data** were collected using a DJI Matrice 600 Pro hexacopter with a Ronin MX gimbal. The imaging was performed with a Resonon Pika L camera (Fig. 1), covering a spectral range of 380–1000 nm across 150 spectral bands. The camera had a 17 mm focal length lens, a 17.6-degree field of view (FOV), and an instantaneous field of view (IFOV) of 0.71 mrad. Data were collected at a constant altitude of 120 m and a flight speed of 1.3 m/s, with flight paths designed using Litchi for DJI Mission Hub to run south to north, minimizing the Bidirectional Reflectance Distribution Function (BRDF) effect. Due to the survey area size, data acquisition occurred over two days: September 6 and 25, 2023, between 12:00 and 14:00 to maximize sunlight and reduce shadows. Flight lines were spaced 15 m apart to ensure a 70% overlap between adjacent hyperspectral cubes. After acquisition, we used Spectronon Pro (v. 3.4.11) for pre-processing, converting radiance to reflectance with calibration target spectra and georectifying hyperspectral cubes using GPS data from the SBG Ellipse IMU and the UAV's onboard GPS. The final hyperspectral mosaic achieved a spatial resolution of 0.2 m.

Multispectral **Sentinel-2 (S-2) data** (Fig. 1), specifically the 2A surface reflectance product, were extracted from the Copernicus Data Space Ecosystem. The selected datasets were captured under cloud-free conditions in

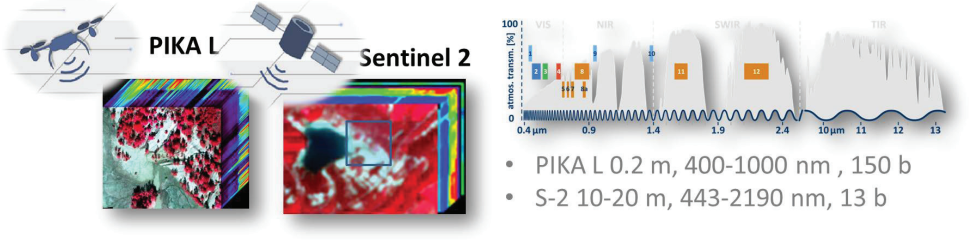

***Figure 1** Schematic illustrating: high spatial resolution hyperspectral data (400-1000 nm) were acquired via UAV using a PIKA L scamera and multispectral Sentinel-2 satellite data, highlighting the spectral range and band positions.*

September 2023, ensuring a rain-free period of at least two days prior to data acquisition.

In-situ and calibration/validation data

Soil and substrate samples were collected as part of long-term research initiatives (2010–2018). Additional samples were gathered during the UAV data acquisition in September 2023. To resolve the sample mineralogy, a Philips X'Pert X-ray Diffractometer (XRD) at the Czech Geological Survey was utilized. The XRD patterns were generated using monochromatic radiation and a graphite secondary monochromator. Random patterns were collected over an angular range of 2° to 70° (2θ), with increments of 0.05° (2θ).

Based on detailed sample mineralogy, we categorized the samples into three distinct classes (Tab. 1, Fig. 2). This classification facilitated precise determination of AMD/pH by monitoring pH stability in Fe sulfates, oxy-hydroxides, and oxides (Swayze *et al.* 2000). We selected different scenarios to train/validate ML models using high spectral and spatial resolution PIKA L data and multispectral 10-m resolution S-2 data. For both datasets, we used sample locations to create regions of interest (ROIs) for training (1 sample per class with established mineralogy) and validation (5 samples per class with established mineralogy) of the Machine Learning (ML) classifications. For high-resolution PIKA L data, mineral classes represented areas of a few meters. In contrast, for S-2 data, we selected more homogenous areas of tens of meters to train the models (Fig. 3).

In addition to mineral samples, we visually identified AMD hotspots using high-resolution orthophotos from Google Earth. While XRD analysis wasn't performed on these hotspots, we confirmed their AMD-generating activity in the field, and this dataset was also used for ML classification validation.

Machine Learning (ML) classifications

Four ML classifications (Shirmard *et al.* 2022) were tested using the ENVI Machine Learning toolbox (v. 5.7) and the Regions of Interest (ROIs) outlined in the previous section were used as labeled data.

Random Forest (RF) is a machine learning technique that uses multiple decision trees trained on different subsets of data. It helps avoid overfitting and works well with large datasets, providing better accuracy and robustness to outliers. However, it can be slow with large forests, may handle categorical variables poorly, and produces larger models. **Extra Trees (ET)**, or Extremely Randomized Trees, is similar to Random Forest but splits nodes randomly without seeking optimal splits. This method samples the entire dataset and is faster than Random Forest, though it shares some disadvantages, such as slow performance with large forests and larger model sizes. **Support Vector Machine (SVM)** is a linear classifier that finds the optimal hyperplane to separate data into classes. It works well when data is linearly separable or nearly so. On the other hand, **Radial Basis Function Support Vector Machine RBF SVM** is a classification algorithm that uses a nonlinear boundary to separate classes, making it effective in high-dimensional spaces. While powerful, it has long training times, limiting its use for large datasets.

Results

The machine learning (ML) classifications were validated using XRD analysis and field-

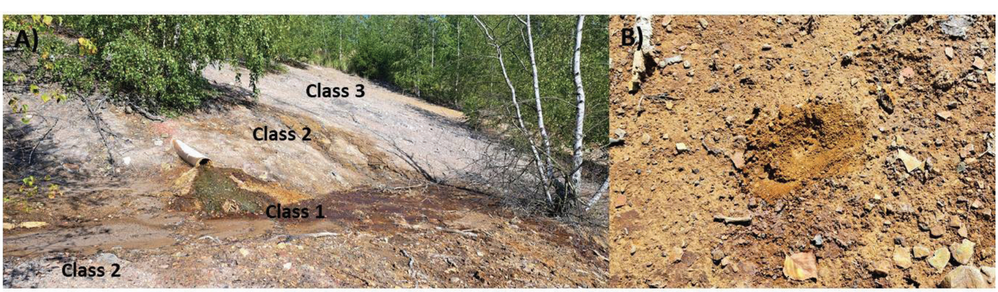

Figure 2 *Lítov dump site - AMD" hot spot" illustrating the field situation and how the mineral classes 1-3 (Tab. 1) look like (A), detailed photo of the class 1 (jarosite-rich) (B).*

Table 1 Mineral classes defined for the Machine Learning classifications and their mineral composition.

Mineral class	AMD minerals XRD	Other minerals XRD
Class 1	jarosite, jarosite>>goethite	gypsum, kaolite, quartz, mica
Class 2	jarosite, goethite/ goethite, jarosite	kaolite, quartz, mica
Class 3	hematite, goethite	kaolite, quartz, mica, lignite

verified AMD hotspots identified visually in Google Earth. Using either PIKA L (Tab. 2, Fig. 4) or S-2 (Tab. 3, Fig. 5) as input, RBF SVM consistently outperformed the other tested ML approaches. With high-resolution PIKA-L data, RBF SVM correctly identified almost all AMD hotspots (6 out of 7 from Google Earth) and accurately matched all mineral classes with XRD-determined mineralogy. The performance of other models, in descending order of accuracy, was: ET, RF, SVM.

Similar results were observed with S-2 data (Tab. 3, Fig. 5), where RBF SVM correctly mapped most AMD hotspots (13 out of 18 from Google Earth) and matched 10 out of 20 mineral compositions determined by XRD. As expected, the precision of AMD mapping decreased with the lower spatial and spectral resolution of the S-2 data. The performance order for other models with S-2 data was: RF, ET, and SVM.

These initial results indicate that RBF SVM excels at detecting AMD discharge and distinguishing mineral mixtures in both multi- and hyperspectral datasets. Its strength likely stems from its ability to effectively define class margins using a nonlinear boundary, making it a robust kernel within the SVM family for high-dimensional data. This allows for accurate differentiation of overlapping classes. While RF and ET classifiers also performed well, they were less accurate in differentiating between jarosite (stable under acidic pH) and goethite/hematite (indicating increasing pH).

Conclusion

The study validated machine learning (ML) classifications for identifying Acid Mine Drainage (AMD) hotspots using XRD analysis and field-verified hotspots from Google Earth. The Radial Basis Function Support Vector Machine (RBF SVM) consistently outperformed other ML approaches, accurately identifying nearly all AMD hotspots with high-resolution PIKA L data (6 out of 7) and matching all mineral classes determined by XRD. For Sentinel-2 (S-2) data, RBF SVM mapped most AMD hotspots (13 out of 18) and matched 10 of 20 mineral compositions from XRD, though its precision decreased due to the lower resolution of S-2 data. The order of performance for other models was Elastic Trees (ET), Random Forest (RF), and Support Vector Machine (SVM) for PIKA L data, and RF, ET, and SVM for S-2 data. uture research will aim to establish connections between AMD mineralogy and specific pH ranges

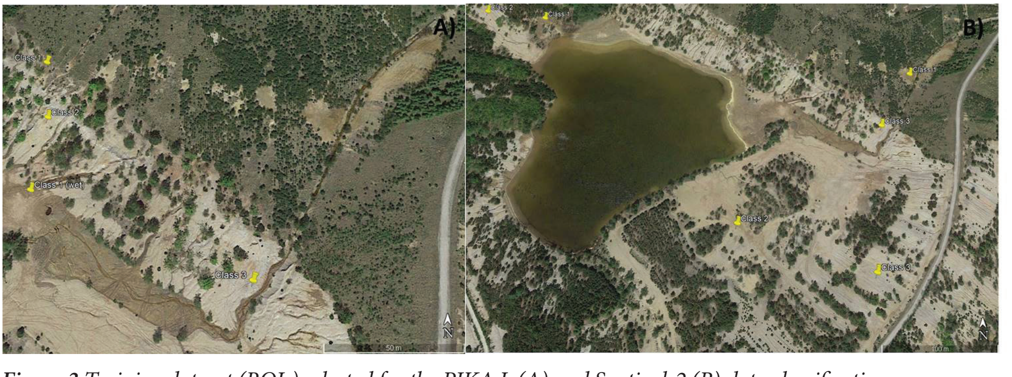

Figure 3 Training dataset (ROIs) selected for the PIKA L (A) and Sentinel-2 (B) data classifications.

Table 2 Validation statistics for the ML classifications (PIKA L data used as the input).

ML classification	Number of correctly/ incorrectly classified AMD hotspots (jarosite-rich class 1) XRD	Number of correctly/ incorrectly classified AMD hotspots Google Earth high-res. imagery	Number of correctly/ incorrectly classified mineral classes 2 and 3 XRD
RBF SVM	6/1	5/0	4/0
SVM	1/6	1/4	1/3
RF	2/5	1/4	4/0
ET	3/4	3/2	4/0

and explore machine learning techniques on extended multi-temporal datasets, broaden training and validation datasets for scalability, and assess model transferability to other locations, such as the Kirki post-mining site in Greece.

Acknowledgement

The presented analysis was conducted under the support of the EC grant MultiMiner. The MultiMiner project is funded by the European Union's Horizon Europe research and innovation actions programme under Grant Agreement No. 10109137474.

References

Flores, H., Lorenz, S., Jackisch, R., Tusa, L., Contreras, I. C., Zimmermann, R., & Gloaguen, R. (2021). UAS-based hyperspectral environmental monitoring of acid mine drainage affected waters. Minerals, 11(2), 182.

Chalkley, R., Crane, R. A., Eyre, M., Hicks, K., Jackson, K. M., & Hudson-Edwards, K. A. (2022). A multi-

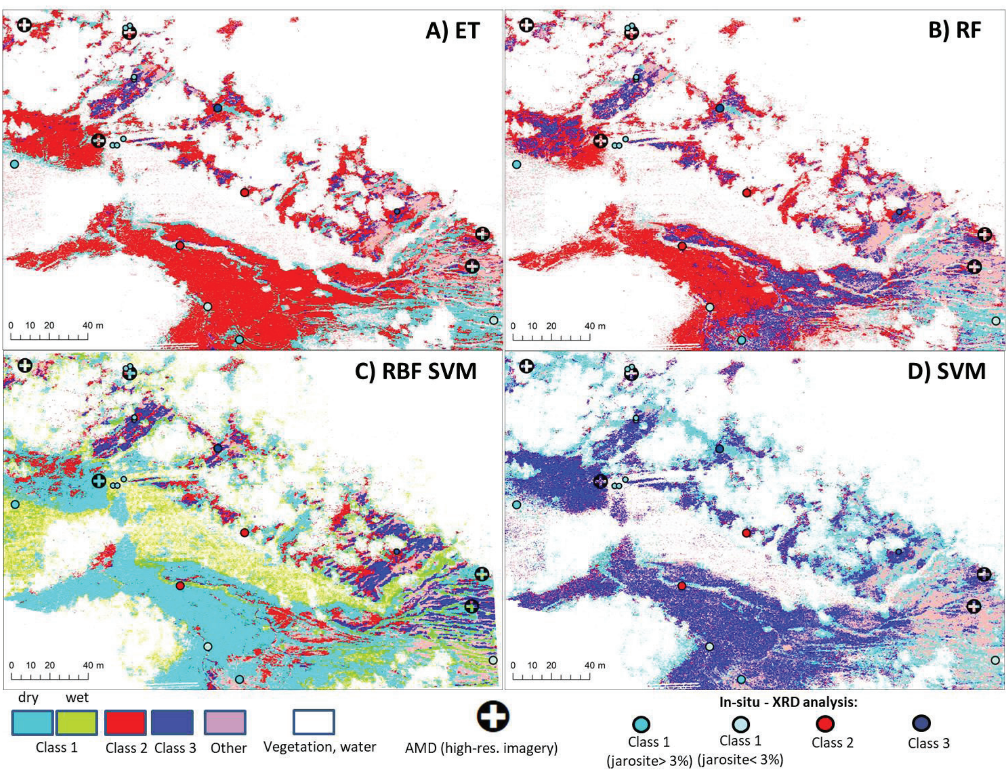

Figure 4 Machine learning classification results for PIKA L data, overlaid with sample locations where XRD analysis was performed for mineralogical determination. The locations of visually identified AMD hotspots from Google Earth imagery are also shown.

Table 3 *Validation statistics for the ML classifications (Sentinel-2 data used as the input).*

ML classification	Number of correctly/ incorrectly classified AMD hotspots (jarosite-rich class 1) XRD	Number of correctly/ incorrectly classified AMD hotspots Google Earth high-res. imagery	Number of correctly/ incorrectly classified mineral classes 2 and 3 XRD
RBF SVM	13/5	5/6	5/4
SVM	11/7	4/7	3/6
RF	12/6	5/6	4/5
ET	10/8	3/8	5/4

scale feasibility study into acid mine drainage (AMD) monitoring using same-day observations. Remote Sensing, 15(1), 76.

Kopačková, V. (2014). Using multiple spectral feature analysis for quantitative pH mapping in a mining environment. *International Journal of Applied Earth Observation and Geoinformation*, 28, 28–42.

Kopačková, V., & Hladíková, L. (2014). Applying spectral unmixing to determine surface water parameters in a mining environment. *Remote Sensing*, 6(11), 11204–11224.

Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. Remote *Sensing of Environment*, 268, 112750.

Swayze, G. A., Smith, K. S., Clark, R. N., Sutley, S. J., Pearson, R. M., Vance, J. S., ... & Roth, S. (2000). Using imaging spectroscopy to map acidic mine waste. *Environmental Science & Technology*, 34(1), 47–54.

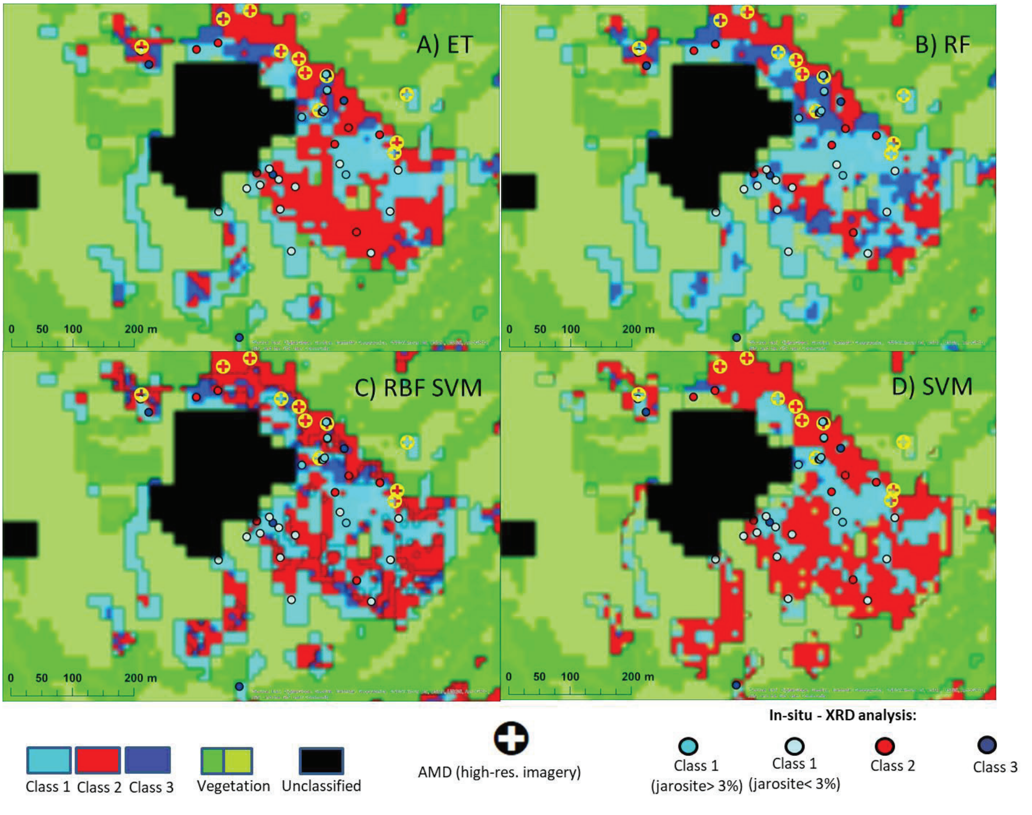

Figure 5 *Machine learning classification results for Sentinel-2 data, overlaid with sample locations where XRD analysis was performed for mineralogical determination. The locations of visually identified AMD hotspots from Google Earth imagery are also shown.*