

Article

Not peer-reviewed version

Convolutional Neural Networks Applied to Antimony Quantification via Reflectance Spectroscopy Using Soils from Northern Portugal: Opportunities and Challenges

[Morgana Carvalho](#)^{*}, [Joana Cardoso-Fernandes](#), [Alexandre Lima](#), [Ana Caudia Teodoro](#)

Posted Date: 26 February 2024

doi: 10.20944/preprints202402.1438.v1

Keywords: Antimony; exploration; reflectance spectroscopy; soil analysis; critical raw materials; deep learning; geology



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Convolutional Neural Networks Applied to Antimony Quantification via Reflectance Spectroscopy Using Soils from Northern Portugal: Opportunities and Challenges

Morgana Carvalho ^{1,2,*}, Joana Cardoso-Fernandes ^{1,2}, Alexandre Lima ^{1,2} and Ana C. Teodoro ^{1,2}

¹ Department of Geosciences, Environment and Spatial Planning, Faculty of Sciences, University of Porto, Rua Campo Alegre, 4169-007 Porto, Portugal; joana.fernandes@fc.up.pt (J.C-F.); allima@fc.up.pt (A.L.); amteodor@fc.up.pt (A.C.T.)

² ICT (Institute of Earth Sciences) – Porto pole (Portugal), Rua Campo Alegre, 4169-007 Porto, Portugal

* Correspondence: morgana.carvalho@fc.up.pt

Abstract: Antimony (Sb) has gained significance as a critical raw material (CRM) within the European Union (EU) due to its strategic importance in various industrial sectors, particularly in the textile industry for flame retardants and as a component of Sb-based semiconductor materials. Moreover, Sb is emerging as a potential alternative for anodes used in lithium-ion batteries, a key element in the Energy transition. This study focused on exploring the feasibility of identifying and quantifying Sb mineralizations through the spectral signature of soils using reflectance spectroscopy, a non-invasive remote sensing technique, and by employing deep learning algorithms such as Convolutional Neural Networks (CNNs). Common signal preprocessing techniques were applied to the spectral data, and the soils were analyzed by Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Despite achieving high R-squared values, the study faces a significant challenge of generalization of the model to new data. Despite the limitations, this study provides valuable insights into potential strategies for future research in this field.

Keywords: Antimony; exploration; reflectance spectroscopy; soil analysis; critical raw materials; deep learning; geology

1. Introduction

Antimony (Sb) is currently considered a critical raw material (CRM) to the European Union (EU), being strategic to its economy, in a scenario where the global market of Sb is dominated by China. The element Sb is included in group 15 (VA) of the periodic table, located in the second long period of the table between tin (Sn) and tellurium (Te). It is classified as a non-metal or metalloid and may exhibit a valence of +5, +3, 0, or -3, with metallic characteristics in the trivalent state [1]. Sb and its mineral sulfides are reported to have been used by humans at least since 4000 B.C. One of its reported uses in more ancient times is as a main ingredient of a black paste, named kohl, used for colouring eyebrows and lining eyes by Egyptians and others in early biblical times [1]. An ornamental vase found at Tello, Chaldea, is reported to be cast Sb and dates to 4000 B.C. The given name to the metal, stibium, is attributed to Pliny the Elder (50 AD), while “antimonium” is reported to be referred to by an Arabian alchemist, Geber, living in the eighth century [2] [1]. A scientific treatise about the element Sb was written by Nicolas Lemery (1645–1715), containing results of his investigations about the properties and different preparations of metal Sb, which was believed to be an important component in the alchemical lore, actuating as a magnet for extracting mercury, a key component for making the Philosopher Stone [3].

Nowadays the main uses of Sb in the EU remain in the textile industry as flame retardants, and as Sb-based semiconductor materials such as lead-acid batteries, lead alloys, catalysts, and stabilisers for plastics, and in the glass and ceramic industry [4]. In the context of the growing demand for electric vehicles, Sb is also being studied as an alternative anode for use in lithium-ion batteries [5].

Today, graphite is mainly employed as an anode in lithium-ion batteries and sodium-ion batteries, although Sb is also considered due to its structure, with a potential for a much better electrical conductor [6].

Several authors have applied machine learning (ML) techniques to detect the presence of heavy metals in soils indirectly through reflectance spectroscopy. Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) were used by Kemper and Sommer [7] to predict the contents of As, Cd, Cu, Fe, Hg, Pb, S, Sb, and Zn, using samples from an area flooded with sludge that resulted from a break of a mine tailings dam in Spain. This accident resulted in a very contrasting mineralogy between the background soil and the contaminated zone. They obtained good predictions for As, Fe, Hg, Pb and Sb. Multiple regression equations were used by Nanni and Dematté [8] to effectively predict Fe_2O_3 and TiO_2 between other soil properties. Partial Least Squares Regression (PLSR) was employed by Cheng et al. [9] to estimate Cd, Pb, As, Cr, Cu and Zn and by Rodríguez-Pérez, et al. [10] for estimating Mn, soil nutrients and other properties such as pH and electrical conductivity. Cheng et al. [9] examined the feasibility of using soil reflectance spectra to estimate the concentrations of the metals in a suburban area of Wuhan City, Hubei Province, China. The concentration of the metals Pb, As, Cr, Cu and Zn in the soil samples were determined by Inductively coupled plasma atomic emission spectroscopy (ICP-AES). They observed how different preprocessing treatments interfere with the results of the prediction model and applied statistical analysis to ascertain the estimation mechanism, focusing on relationships between soil reflectance spectra and concentrations of Soil Organic Matter (SOM), Fe, and heavy metals (Cd, Pb, As, Cr, Cu, and Zn). They concluded that employing Savitzky-Golay spectral pre-treatment yielded favorable PLSR models, however, additional studies were needed to establish internal relationships between the heavy metal concentrations and spectrally active elements, as the authors did not achieve satisfactory results. Rodríguez-Pérez et al. [10] were able to obtain a good performance only for phosphorus, pH and electrical conductivity. Pyo et al. [11] used a CNN and compared it to ANN and Random Forest Regression (RFR) to estimate As, Cu and Pb soil samples taken from a mining area, located in the Geum River watershed of South Korea, and obtained the best results with the CNN model, but also achieving reasonable soil heavy metal estimation accuracy with the other machine learning models. RFR was used by Guo et al. [12] to successfully infer the Zn and Ni concentrations based on the relationship between heavy metals with soil criteria and clay. RFR together with PLSR and Support Vector Machine (SVM) were used to predict Mn, Cu, Zn, Pb, Cr and Ni contents, obtaining the best results with RFR.

CNNs are well-established in various domains, including object detection, image classification, and spectral analysis [13]. By leveraging sparse local connections and weight sharing, CNNs have proven to be effective in learning and extracting local and abstract features from raw spectral data. By stacking multiple convolutional and pooling layers, the CNN model can efficiently capture intricate patterns within the data, making it well-suited for soil content prediction tasks [13, 14].

The objective of this study was to verify the possibility of identifying Sb mineralization through the spectral signature of soils with the application of a CNN, a deep learning technique. The major advantage of developing ML and deep learning (DL) methods to quantify and qualify heavy metals in soils is the possibility of analyzing large-scale zones faster, avoiding the high cost and time demanded that is implicit in the traditional approach of geochemical analysis. Reflectance spectroscopy is a remote sensing technique, non-invasive, capable of identifying targets for mineral exploration reducing costs and avoiding environmental impacts. The soils samples for this study were obtained and analyzed in previous studies for characterizing the Sb distribution in the former mining areas of Ribeiro da Serra and Tapada, in Northern Portugal. The CNN model shows promising results, but in this study the overfitting of the model couldn't be avoided. Despite not having achieved good results for the Sb predictions, this study provides insights about which strategies could be incorporated into future studies.

2 Background

2.1 Study area

The study area encompasses the two former Sb-Au mining concessions of Ribeiro da Serra and Tapada, located in northern Portugal, approximately 30 km East from the city of Porto (Figure 1). The Ribeiro da Serra and Tapada mines were opened in 1880 and 1881, respectively, producing thousands of tonnes of antimonite concentrates annually for exportation [15, 16]. The exploitation of Sb hit its peak in the 19th century, but in the first years of the 20th century, the competition with the Asiatic Countries led to the closure of the Portuguese Sb mines. During the Second World War, there was an increase in mining activity, and since the 1960s some prospecting campaigns and reconnaissance studies were executed [17]. Nowadays the mining structures are abandoned (Figure 2) and many waste piles and tailings remain in the zone.

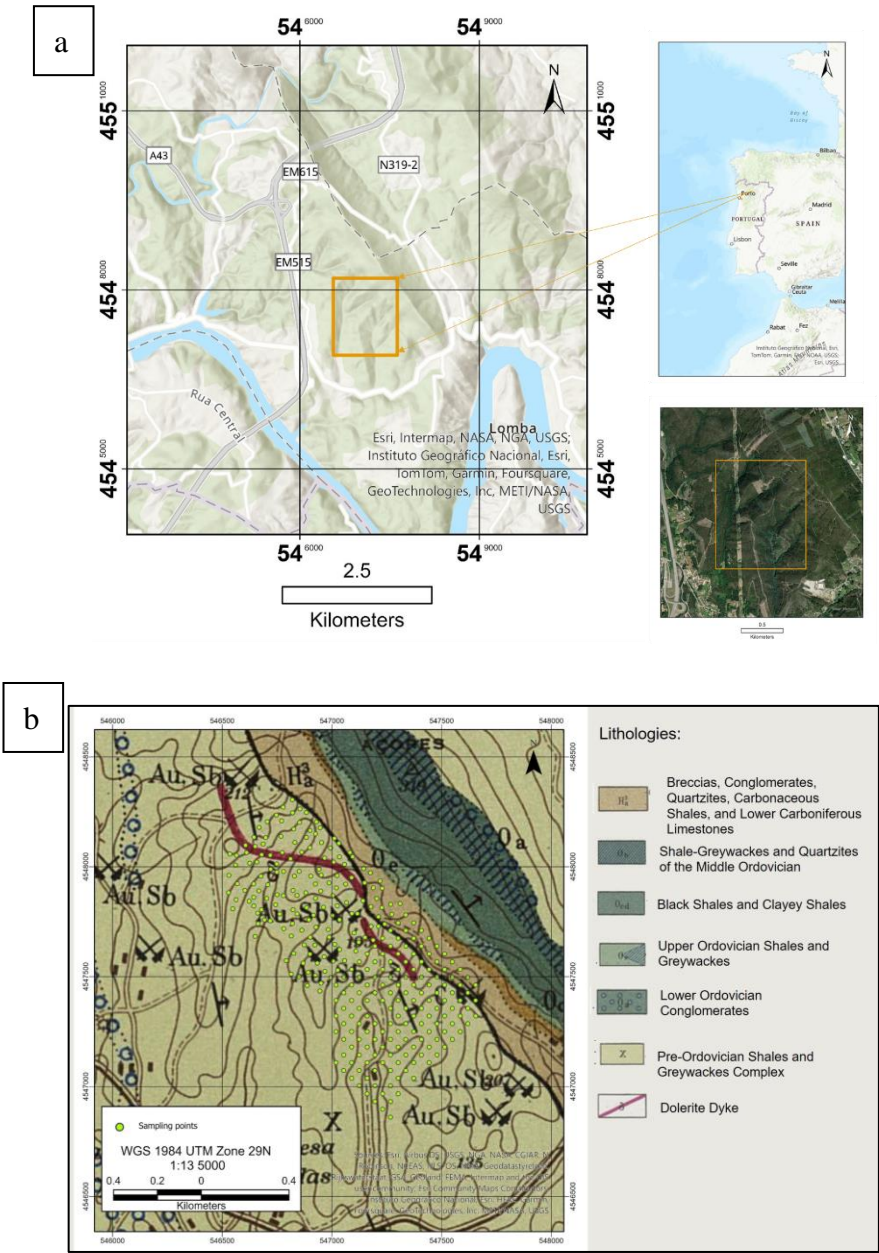


Figure 1. (a) Sampling points (green dots) and geology of the study area.

Geologically, those Sb deposits are located on the western flank of a Variscan structure, the Valongo Anticline, in the Dúrico-Beirão Mining District, situated within the Iberian Central Zone [16, 18-20]. The lithostratigraphic succession consists of Cambrian/Pre-Cambrian (pre-Ordovician) rocks with very low-grade metamorphism; Ordovician and Carboniferous sequences, composed of schists and some quartzites, and the Upper Carboniferous formation comprising breccias, conglomerates, and intercalated quartzites. The lithologies vary from east to west, with ages corresponding to Lower Ordovician, Middle and Upper Ordovician, Lower Carboniferous, and pre-Ordovician.

The Sb mineralization occurs in low volume in discontinuous quartz veins that are mainly hosted in Silurian schists and greywackes [20]. The quartz veins have a hydrothermal nature and are associated with Variscan granitic intrusions, or due to fluid mixing of CO₂-rich metamorphic fluids by surface-derived H₂O–NaCl fluids [15]. The dominant directions of the country rocks range from N to NW dipping to W. The most productive veins occur in the EW direction dipping to N (Tapada) and NS dipping to W (Ribeiro da Serra).



Figure 2. Ribeiro da Serra mine infrastructures picture in the late 19th century [21] (a) and nowadays ruins (b).

2.2. Spectral reflectance

Reflectance spectroscopy offers a means to extract multiple soil properties, both direct and indirect, as well as metal contents [22]. The abundant data generated by soil spectroscopy, whether in the form of point measurements or images, necessitates the implementation of data-modelling procedures.

Materials may reflect or absorb electromagnetic radiation at varying wavelengths, governed by factors such as surface absorption, emissivity, and reflectance characteristics. The spectral range employed for soil reflectance analysis encompasses the Visible and Near-Infrared to Short-Wave Infrared (VNIR-SWIR) region, spanning from 400 to 2,500 nm. This range is further divided into two sub-ranges: VNIR (400 to 1,100 nm) and SWIR (1,100 to 2,500 nm). The interactions between light and matter are intricately tied to the wavelength. Though pure metals don't exhibit absorption within the VNIR-SWIR region, their presence can be indirectly detected through associations with organic matter (OM), interactions with compounds like hydroxides, sulphides, carbonates, or oxides that manifest detectable properties, or adsorption to light-absorbing clays [22].

2.2. Convolutional Neural Networks

The field of geosciences is having a slow but crescent incorporation of ML and DL techniques. ML data analysis methods can automate the creation of analytical models and perform tasks such as classification, regression, or clustering [23-26] that are useful in the geosciences. The improvement of the model is given by its training with a data set that will condition the results, and evaluation metrics

are implemented to quantify the performance of the model with the given data set. Additional samples of data often can improve the model performance and the aim is to obtain a good model that can generalize to new data [23]. The difficulty in the availability of labelled data is exactly what makes the implementation of ML and DL remote sensing applications to geosciences a challenging task. Frequent problems of this nature are the limitation of possibilities in the data collection; the large number of physical variables associated with a limited number of samples; and the difficulty of obtaining high-quality measurements of several geoscience variables, that can only be taken by overly expensive or time-consuming techniques[27]. The heterogeneity of the data and the multi-resolution are often associated with the already challenging multiconnected nature of the different processes in geosciences [27]. Despite those challenges, the capacity for ML and DL methods to actuate in geosciences tasks that were only possible to be executed with dispendious human work, and even in tasks that were not possible to be executed before, make their implementation increasingly attractive.

The capacity of CNN to capture nonlinear behaviors makes it suitable for geological problems. CNN demonstrated powerful abstraction capabilities in the field of geosciences, with the first applications in the field focused on seismic interpretation, more recently being incorporated in broader geosciences applications such as global water storage modeling, landslide prediction, and earthquake arrival time picking [14] [23]. The constitution of CNNs primarily involves three types of layers: convolutional layers, pooling layers, and fully connected layers [13]:

1. Convolutional Layer: In this layer, input features are convolved with learnable kernels, generating various output feature maps. Each kernel has a fixed length and slides over the input feature map with a stride of 1, performing convolution with local regions where the kernel overlaps the input feature map. Non-linear activation functions are then applied to the convolution results to produce output feature maps. This layer's key advantage lies in its ability to learn local patterns while significantly reducing the number of model parameters through weight sharing.

2. Pooling Layer: After convolution, the pooling layer is employed to progressively reduce the spatial size of the output feature maps generated by the convolutional layer.

3. Fully Connected Layer: The fully connected layer comprises neurons that connect to all activations extracted from the convolutional and pooling layers. This layer plays a crucial role in generating the final output results.

3. Materials and Methods

3.1. Soil collection and preparation

The present study benefited from the soil samples collected in the scope of the AUREOLE project-ERA-MIN/0005/2018 (<https://aureole.brgm.fr/>; access on 25 January 2024). The soil sampling campaigns were carried out covering the area where the underground works in Ribeiro da Serra and Tapada took place and the surrounding areas. The soil sampling aimed to test the mineralization distribution and identify possible new mineralized structures and soil contamination distribution. The first soil sampling campaign was carried out in the Ribeiro da Serra Sb-Au mine zone in 2021, with 157 samples collected, while the second soil sampling campaign, in the Tapada mine zone took place in 2022, with 152 samples collected. The sampling campaigns follow a plan on a grid of 50x50 meters in the area that with an orientation according to the mining structures. Soils were collected from horizon B and horizon C when there was no horizon B which is relatively common in the study area. From the totality of samples, 54 samples from Ribeiro da Serra and 53 samples from Tapada were sent to Bureau Veritas laboratory in Vancouver (Canada) to be analyzed by ICP-MS (Inductively coupled plasma mass spectrometry analysis). Before sending the samples to ICP-MS analysis, the samples were dried in a muffle furnace or at least 48 hours and at a temperature of 55 °C. With the samples dry, after rifling the soil samples were grounded to a size < 200 µm. From the samples sent to ICP-MS analysis, only 99 had their spectral signature collected, due to some samples no longer being available.

3.2. Soil reflectance measuring

The raw soil, previously dried with gravel and pieces of plants removed, was used for taking the reflectance measurements. The soil was spread in a watch glass above a black surface and five different points of the sample were measured, resulting in five spectra per sample. Each spectrum collected is a result of an average of several measurements. For this work an average of 5 measurements was used, each measurement resulted in 40 scans.

The FieldSpec 4 standard resolution spectroradiometer equipment (ASD Inc., Boulder, CO, USA) was used to collect the spectral data. The proceeding included “heating” the equipment for 30 minutes before starting to measure. The spectroradiometer has three sensors, one VNIR and two SWIR sensors. The time of “heating” is necessary to ensure that the three sensors are at the same temperature [28]. Normalization with a perfect albidum (white plate) also is needed every time a new measurement project is started and every two hours of work. The software used was Indo Pro [28].

3.2. Spectral preprocessing

The first step was removing from the dataset samples with Sb content above 1000 ppm, that were considered outliers, for being mostly associated with contaminated areas that don't represent the Sb concentrations found related with the natural occurrence of mineralization (see section 3.1), common spectral preprocessing was implemented to improve the results obtained by eliminating noise and highlighting spectral features. The wavelengths before 400 nm were removed due to the excessive noise in this zone of the spectra. The pre-processing steps included: convert the reflectance to absorption; remove the continuum; smoothing the signal and calculate the first and second derivatives; convert the waveform spectrogram.

The order of the spectral preprocessing is given in the diagram (Figure 3).

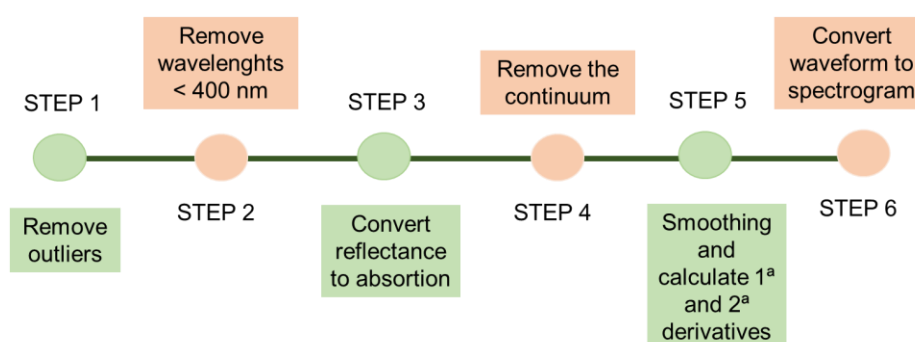


Figure 3. Preprocessing steps followed in this study.

3.2.1. Continuum removal using convex hull

Continuum removal is a technique that allows the extraction of characteristic absorption bands on the reflectance spectrum curves [29]. The convex hull forms a polygon connecting the outermost points within the sample while ensuring that all the internal angles of this polygon are less than 180 degrees, forming the smallest convex shape that encloses all the points in a given set [30]. The continuum-removed data was obtained using a Python script by [28].

3.2.2. Reflectance to absorption, Smoothing the data and calculate the First and Second derivatives

The absorption was calculated from the data, after the continuum removal, using a function implemented in Python, as mentioned in [7], based on reflectance using the logarithmic relationship. After the conversion to absorption, the data was smoothed by applying the Savitzky-Golay filter from the `savgol_filter` function in the `scipy.signal` module in Python. To smooth the signal data, the Savitzky-Golay filter calculates a polynomial fit of each window based on polynomial degree and window size. The obtained smoothed data is used to calculate the First and Second derivatives by applying the `np.gradient` function from the NumPy library.

3.2.3. Convert waveform to spectrogram

This step consists on the application of the Short-Time Fourier Transform (STFT) to the equal_length tensor using tf.signal.stft. A new dimension is added to the spectrogram tensor using spectrogram [..., tf.newaxis]. This dimension is included to the data to make the spectrogram suitable as input with convolution layers in the CNN. This step is essential to convert the two-dimensional data in a three-dimensional data, such structure is required for the application of a CNN. This methodology can be replicated by the application of the Python script available in the supplementary materials (Code S1).

3.3. Application of Convolutional Neural Network

In the present work, to deal with limited computational capacity, a MobiletNet model was used. MobileNets are a class of highly efficient CNN models, built upon a streamlined architecture that leverages depth wise separable convolutions, being a deep neural network with significantly reduced computational demand [31]. The model was implemented using the open-source libraries of TensorFlow and Sklearn and is available in the supplementary materials (Code S2). The model was tested for Sb and other elements, As, Pb, Mn, Zn, that also are not directly detected in the SWIR-VNIR spectral range and have exhibit different Person’s correlation with Sb in the study area (Table 2). The metrics for evaluate the model performance consist in the analyzes of the R² and RMSE obtained.

Table 2. Person’s correlation for the selected elements in the study area.

Element	Sb	As	Pb	Mn	Zn
Sb	1	-	-	-	-
As	0.9	1	-	-	-
Pb	0.63	0.73	1	-	-
Mn	0.15	0.28	0.42	1	-
Zn	0.25	0.41	0.72	0.66	1

3. Results

3.1. ICP-MS analysis

The results obtained from the ICP-MS analysis are shown in Table A1 with the indication of which samples were discarded from the training set. The distribution of Sb concentrations in the study area are depicted in Figure 4.

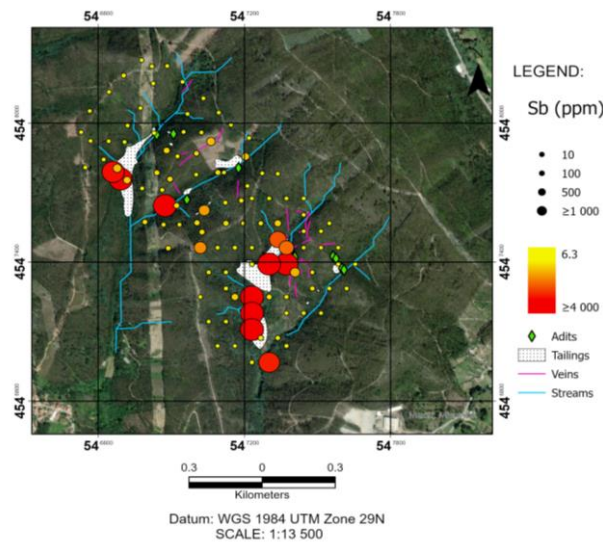


Figure 4. Distribution of Sb concentrations in the study area, and the position of old mining adits, the known veins, tailings and local streams.

The higher concentrations of Sb, the values above 1000 ppm, which were 10% of the totality of the samples, are related to the soils collected in tailings, near the tailings, or in the streamlines. Mostly the values of Sb are between 10 and 100 ppm and are not related to the adits and some are related to veins or streams. Values between 500 and 1000 ppm are related to known veins and others can be associated with the presence of unknown veins. Also, some of those soils have proximity to the tailings and higher values of Sb can be influenced by contamination left by the mining works that took place in the last century.

3.2. Preprocessing and Deep Learning Model Results

Removing the outliers based on the Sb concentrations, training the model with only the samples that contained up to 1000 ppm of Sb content, had a positive impact in the performance of the model, leading to higher values for R². Similarly, the application of the preprocessing steps and removal of the wavelengths before 400 nm successfully improve the results. Oppositely, removing other portions of the spectra (1500-2400) did not improve the model performance. Regarding the preprocessing steps, the best results were obtained by applying the first derivative, while the second derivative did not improve the results.

Only the results of the best combination of preprocessing methods are presented. These results correspond to the signal used has input using the wavelengths between 400 and 2400 nm, using the reflectance converted to absorption; spectra after removing the continuum; signal smoothing and calculation of the first derivative. The input was this processed signal converted to waveform spectrogram. Training the model using multiple elements instead of a single element was tested for making the predictions, but it did not improve the results or reduce the overfitting, so, only the results for single elements are presented (Table 3).

Table 3. Elements and R², RMSE for train and validation and the number of epochs of training.

Element	R ²	RMSE train	RMSE validation	Training epochs
Sb	0.7	0.0014	173	1000
As	0.96	0.01	46	1000
Pb	0.83	0.04	20	750
Mn	0.93	0.0006	41	600
Zn	0.78	0.0002	18	1000

In the results obtained, despite achieving relatively high R^2 values, there is a notable issue of overfitting. The model learned to predict the values of Sb for the training set, but with a high validation error (Figure 5).

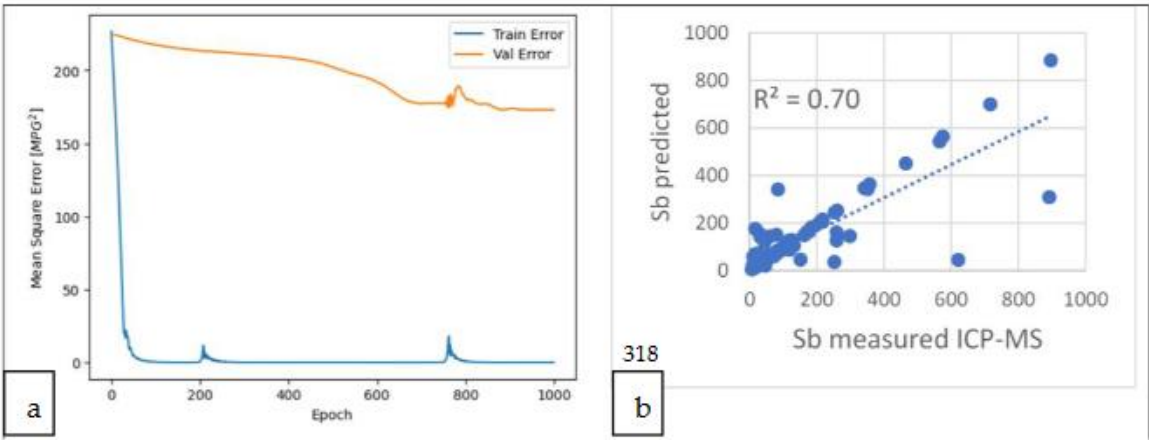
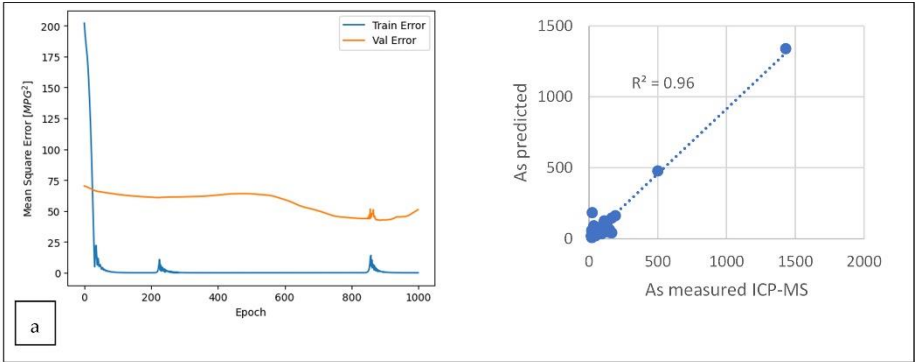


Figure 5. (a) train error versus validation error by epoch for Sb. (b) R^2 for Sb predicted and measured.

Overfitting occurs when the model learns the training data too well, capturing patterns specific to the training set but failing to generalize well to new, unseen data [32]. In this context, despite the promising R^2 values obtained for the elements As, Pb, Mn, and Zn, the disparity between the RMSE values for the validation set is considerably large (Figure 6).



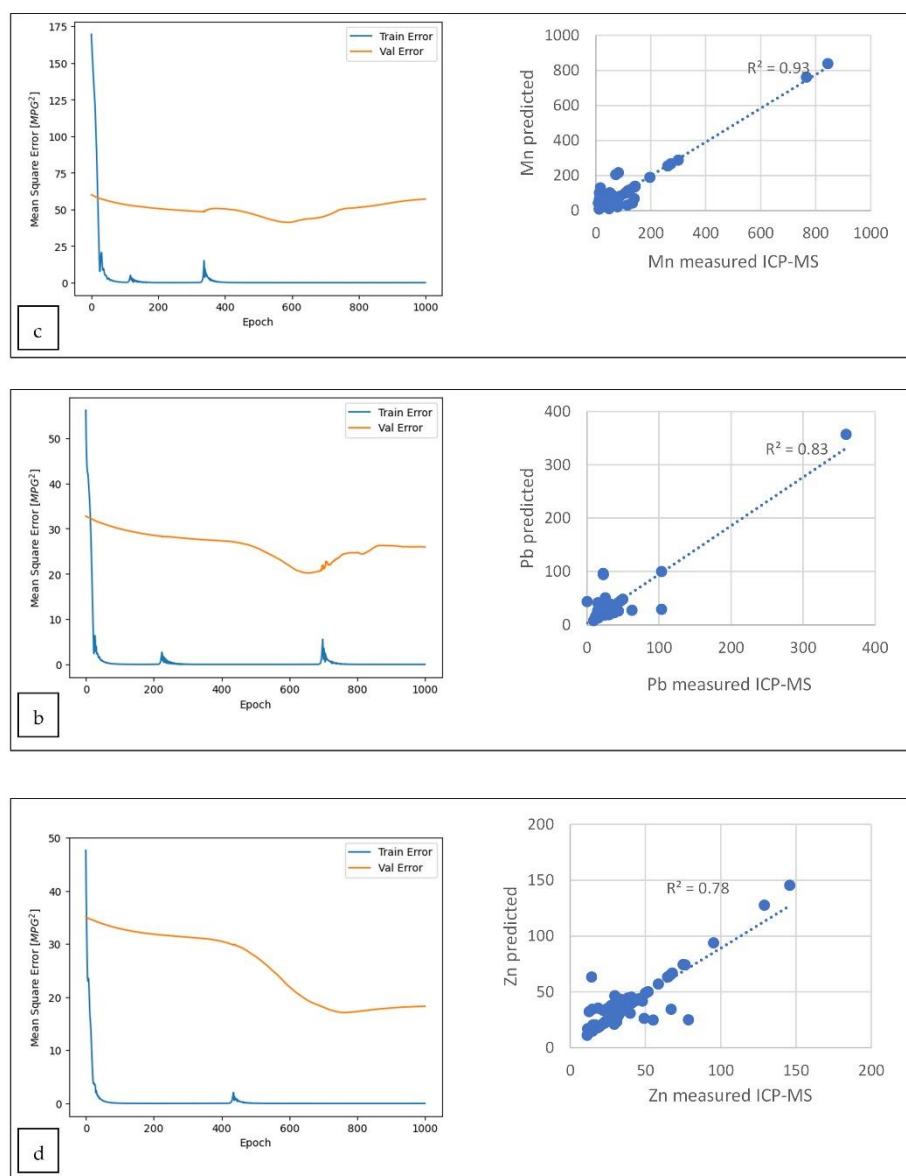


Figure 6. (a) train error versus validation error by epoch and R^2 for a) As, b) Pb, c) Mn, d) Zn.

The observation of the discrepancy in the RMSE between train and validation, while the R^2 values for the training set are notably high, indicating a good fit to the training data, signifies that the model couldn't achieve a good generalization performance. However, some elements experiment better results for the generalization, namely As and Mn. The overfitting tendency is more pronounced for Sb and Zn. We can observe in the graphics in Figure 7 that, for all the elements, the model learns very well how to predict the training set in the first epochs, while the validation error stays in a plateau. It is worth mention that while the validation error may appear to be in a tendency to decrease, tests executed with a few thousand of epochs more show that there is no improvement in the predictions. Those results may indicate that there is a limit in the generalization that the model can reach with the present data set.

4. Discussion

Despite achieving relatively good R^2 values, the presence of significant overfitting weakens the reliability and generalizability of the model's predictions in the present study. Additionally, the observation that, incorporating additional elements into the model training process did not lead to

improved results or mitigate overfitting, further underscores the challenge of addressing this issue. This suggests that simply increasing the complexity of the model or incorporating more features does not necessarily yield better performance and may exacerbate overfitting instead. Kemper and Sommer [7] to resolve the overfitting issue, used a methodology to degrade the spectra considering the band center and the full-width half-maximum. The band center refers to the central wavelength or position of a spectral feature or band of interest. This approach was not possible to replicate in the current study, because the band center for the target element is unknown. Also, in their study area, they have a big contrast between the soil of the region, and the contaminated soil, which was from a mine dump and had a high concentration of heavy metals. In the present study, there is no big contrast between the soils that present Sb and the soils that don't, and the Sb contents are relatively discrete. Wu, et al. [33] found the correlation with total Fe, active and residual, was a major predictive mechanism for heavy metals in soils. Also, OM and clay have a correlation. The soil analysis in the current study didn't include those proprieties, which can be a way to obtain better results.

As the soil sampling campaigns executed focused in capturing the general distribution of the Sb, they do not capture the progressive increment in Sb in mineralized zones. Moreover, many soil samples capturing anomalous values of Sb are sourced from the mine tailings existent in the region, and their properties may not be representative. Another sampling methodology, focusing on the soils near the known Sb veins and on the progressive contents of Sb in the soils associated with the veins could be more appropriate for this study and can work as a solution for the overfitting in the CNN model. And, also, obtaining organic matter and clay data from the soils can be a different approach that can help to better understand the features in the spectral signature of soils containing Sb that can be employed for its identification.

5. Conclusions

This study found varying concentrations of Sb in the sampled area, with the higher values of Sb influenced by the historical mining activities and potential contamination. The implementation of a CNN with low computational demand, MobileNet model, for predicting Sb values shows promising results with a good fit for the training data, but with issues in to generalize to new data. However, challenges emerged regarding its ability to generalize to new data. Notably, preprocessing steps remain essential for enhancing model performance. Alternative sampling methodologies and the increment in the dataset available, as the incorporation of another soil proprieties such as OM and clay into the analysis could provide more insides to the topic. This study provides insights into the application of deep learning models to predict Sb concentrations using spectral data, while there are still challenges to overcome.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Code S1: waveform to spectrogram; Code S2: Convolutional neural network.

Author Contributions: Conceptualization, M.C., J.C-F. and A.L.; methodology, M.C. J.C-F. and A.C.T.; software, M.C.; validation, M.C.; formal analysis, M.C.; investigation, M.C.; resources, A.L.; data curation, M.C.; writing—original draft preparation, M.C.; writing—review and editing, J.C-F, A.C.T., A.L.; visualization, M.C.; supervision, J.C-F, A.C.T. and A.L.; funding acquisition, A.L. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the support provided by Portuguese National Funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P. (Portugal) projects UIDB/04683/2020 and UIDP/04683/2020 (Institute of Earth Sciences). Additionally, the authors express their gratitude to the Aureole project (10.54499/ERA-MIN/0005/2018) for providing the samples used in this study.

Data Availability Statement: Geochemical analysis results are provided within the study in Appendix A. All the Python code developed in this study is freely available as Supplementary Material. Spectra continuum removal and absorption extraction were accomplished using a Python routine publicly available at <https://www.mdpi.com/2306-5729/6/3/33/s1> (access on 19 February 2024), © Copyright 2021 by Cardoso-Fernandes, J.; Silva, J.; Dias, F.; Lima, A.; Teodoro, A.C.; Barrès, O.; Cauzid, J.; Perrotta, M.; Roda-Robles, E.; and

Ribeiro, M.A., under a Creative Commons Attribution (CCBY) license, based on the PySptools open-source Python library, © Copyright 2013–2018, Chris tian Therien, licensed under an Apache License Version 2.0. and available on GitHub repository <https://github.com/ctherien> access on 19 February 2024). Spectral data used in this paper is available in csv format at: <https://doi.org/10.5281/zenodo.10684797>.

Acknowledgments: Special appreciation is extended to Giulia Resta and Ana de Carvalho for their invaluable contributions to the field sampling campaigns and the preparation of soil samples for analysis.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Samples analysed by ICP-MS from Ribeiro da Serra (RSXXX) and Tapada (TPXXX) mining areas and the values obtained for the elements used in this study.

Sampl e	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)	Sampl e	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)
RS001	13.38	25	33.49	13	12.7	TP004	298.5 5	146.9	38.91	36	31.9
RS004	12.79	24.5	17.05	41	25.1	TP007	80.07	67.9	21.95	29	42.9
RS007	10.71	18.7	22.23	68	24.5	TP008	184.6 1	189.9	35.35	24	17.8
RS013	23.4	24.6	22.19	56	22.7	TP009	31.39	46	18.38	29	34.6
RS015	14.55	16	28.89	60	49.1	TP011	27.01	32.5	30.96	299	64.4
RS017	9.95	17.4	23.97	44	23.7	TP013	251	167.4	24.41	17	30.8
RS019	9.5	21.3	21.08	22	14.7	TP016	357.0 4	50.7	26.71	37	29.1
RS021	33.94	22.6	28.76	53	36.5	TP017	1446	467.8	35.95	376	47.2
RS023	30.32	21	26.2	68	41.1	TP027	52.6	70.3	25.98	23	19.7
RS027	16.87	17.9	24.53	51	31.1	TP033	68.45	166.5	27.97	59	40.7
RS029	9.54	14.8	20.64	34	29.1	TP035	715.4 9	78	49.92	142	68
RS031	25.73	20.3	20.93	57	29.8	TP038	116.3 3	50	25.42	27	22.1
RS034	47.03	112.9	19.54	40	24.5	TP040	258.4	74.5	62.37	63	26.9
RS037	575.47	1431. 2	47.29	76	49.8	TP044	259.1 8	65.7	42.6	844	146
RS040	59.42	28.6	33.97	261	50.9	TP048	14.87	23.4	23.87	229	54.1
RS045	19.46	47.8	17.06	13	15	TP049	207.6 5	30.9	25.6	27	19.7
RS049	162.61	111	24.82	35	28.3	TP051*	>4000	571.8	325.79	35	26.4
RS051*	2653	225	31.82	31	26.1	TP055	50.4	27	25.04	51	31.3
RS052	2215	120.9	40.35	95	38.6	TP064	103.8 2	34.8	33.38	106	65.3
RS055	50.68	21.2	12.64	15	23.6	TP065	352.4 7	71	22.64	26	26
RS057	30.04	20.8	20.98	73	39.7	TP068	217.9 4	47.9	20.33	37	31
RS061	82.11	30.8	25.48	24	25.9	TP070	891.0 9	99.7	25.65	132	67
RS066*	>4000	501.9	197.0 1	28	17.5	TP072	6.3	16.4	29.52	57	33
RS067*	1103	129.6	28.5	18	21	TP075	50.02	51.3	20.59	13	22.8

RS069	170.9	29.6	23.63	36	26.3	TP077	258.2 2	27.4	22.34	26	29.4
RS073	342.44	84.8	34.84	27	36.1	TP079	85.29	33.8	22.7	37	30.1
RS076	59.27	85.2	20.77	17	14	TP082	31.88	22.8	14.67	51	32.7
RS078*	>4000	680.5	171.8 9	143	51.8	TP084	22.03	23.9	32.51	82	40.8
RS084	59.11	21.1	17.59	38	22.3	TP087	619.7 1	92.3	103.39	78	38
RS088	89.74	34.8	20.79	30	22.8	TP091	464.2 4	130.1	17.55	34	22.6
RS090	101.96	76.7	22.39	48	25	TP093	38.45	19	9.26	767	128.8
RS094	79	91.5	24.93	29	26.5	TP094	13.69	21.6	29.53	72	47.9
RS096	119.56	53.9	25.81	27	19.4	TP099*	>4000	1208. 1	1040.1 4	143	110.8
RS100	127.62	43.8	18.54	35	28.3	TP106	45.26	38.8	22.23	119	45.5
RS106	252.69	63.4	103.5 1	139	95.3	TP109*	3786	240.8	190.75	387	55.4
RS108	61.45	72	30.65	28	28.9	TP111	895	499	29.82	197	66.7
RS112*	1712	313.1	125.7 6	27	43.9	TP115	93.15	37.9	18.36	271	76.2
RS115	118.07	147	25.04	49	33.4	TP122	33.43	21.6	24.65	124	58.4
RS118*	>4000	966.6	449.0 3	442	74.9	TP125	44.84	19.4	19.12	44	22
RS126	131.8	48.8	23.47	22	25.5	TP132	10.99	23.4	28.35	139	55.2
RS128	99.13	33.6	15.16	28	14.5	TP133	38.25	25.5	26.92	61	33.9
RS129	565.34	81.6	359.7 5	88	26.7	TP136	33.92	42	27.22	72	43.5
RS131*	>4000	895.5	228.8 6	90	34.7	TP138	48.68	17.7	20.54	41	42.9
RS134	151.98	91.1	37.56	78	30.1	TP140	16.53	25.1	19.3	17	18.1
RS135	103.69	60.4	30.77	35	23.3	TP147	26.18	30.8	43.43	116	78.4
RS143*	>4000	671.9	221.6 8	163	41.1	TP154	17.61	23.3	16.89	28	34.5
RS144	217.65	47.3	25.37	28	20.1	TP156	13.44	27	15.95	33	27.4
RS148	48.3	36.2	16	11	11.1	TP159	13.91	17.2	36.37	143	74.9
RS151	45.82	37.6	12.26	9	16.6	TP167	11.39	20.3	12.68	15	20.6
RS156*	3500	361.7	50.46	46	28.6	TP173	6.53	20.8	21.16	115	51.8
RS159	69.45	31	15.37	18	11.3	TP175	16.54	17.9	15.22	16	18.2
RS162	68.91	28.6	26.37	55	34.2	TP177	176.9 8	44.5	18.68	23	50.2
RS164	79.33	24.6	19.59	57	31.3	TP179	10.66	50.7	15.55	17	23.3
RS169	72.12	20.7	15.76	21	24.8						

* Samples excluded of the training set.

References

1. Li, T., Archer, G.F. and Carapella, S.C., Jr. Antimony and Antimony Alloys. In *Kirk-Othmer Encyclopedia of Chemical Technology*, John Wiley & Sons, Inc **2000**; pp 1-15.

2. Butterman, W.; Hilliard, H. Mineral commodity profiles. *Selenium. Rapport US Department of the Interior US Geological Survey*; Professional Paper 1802–Q, Reston, Virginia, USA, **2004**, 1-20.

3. Wisniak, J. Nicolas Lémery. *Revista CENIC. Ciencias Químicas* **2005**, 36 (2), 123-130.

4. European Commission; Directorate-General for Internal Market Industry Entrepreneurship and SMES; Grohol, M.; Veeh, C. Study on the critical raw materials for the EU 2023 – Final report. Publications Office of the European Union: 2023.

5. Moolayadukkam, S.; Bopaiah, K. A.; Parakkandy, P. K.; Thomas, S. Antimony (Sb)-Based Anodes for Lithium-Ion Batteries: Recent Advances. *Condensed Matter* **2022**, *7* (1). DOI: 10.3390/condmat7010027.
6. He, J.; Wei, Y.; Zhai, T.; Li, H. Antimony-based materials as promising anodes for rechargeable lithium-ion and sodium-ion batteries. *Materials Chemistry Frontiers* **2018**, *2* (3), 437-455. DOI: 10.1039/c7qm00480j.
7. Kemper, T.; Sommer, S. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental science & technology* **2002**, *36* (12), 2742-2747. DOI: 10.1021/es015747j.
8. Nanni, M. R.; Demattê, J. A. M. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Science Society of America Journal* **2006**, *70* (2), 393-407. DOI: 10.2136/sssaj2003.0285.
9. Cheng, H.; Shen, R.; Chen, Y.; Wan, Q.; Shi, T.; Wang, J.; Wan, Y.; Hong, Y.; Li, X. Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy. *Geoderma* **2019**, *336*, 59-67. DOI: 10.1016/j.geoderma.2018.08.010.
10. Rodríguez-Pérez, J. R.; Marcelo, V.; Pereira-Obaya, D.; García-Fernández, M.; Sanz-Ablanedo, E. Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy* **2021**, *11* (10). DOI: 10.3390/agronomy11101895.
11. Pyo, J.; Hong, S. M.; Kwon, Y. S.; Kim, M. S.; Cho, K. H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. *Sci Total Environ* **2020**, *741*, 140162. DOI: 10.1016/j.scitotenv.2020.140162.
12. Guo, B.; Guo, X.; Zhang, B.; Suo, L.; Bai, H.; Luo, P. Using a Two-Stage Scheme to Map Toxic Metal Distributions Based on GF-5 Satellite Hyperspectral Images at a Northern Chinese Opencast Coal Mine. *Remote Sensing* **2022**, *14* (22). DOI: 10.3390/rs14225804.
13. Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis-NIR spectroscopy. *Geoderma* **2020**, *380*, 114616. DOI: <https://doi.org/10.1016/j.geoderma.2020.114616>.
14. Mamalakis, A.; Barnes, E. A.; Ebert-Uphoff, I. Investigating the Fidelity of Explainable Artificial Intelligence Methods for Applications of Convolutional Neural Networks in Geoscience. *Artificial Intelligence for the Earth Systems* **2022**, *1* (4). DOI: 10.1175/aies-d-22-0012.1.
15. Neiva, A. M. R.; Andráš, P.; Ramos, J. M. F. Antimony quartz and antimony-gold quartz veins from northern Portugal. *Ore Geology Reviews* **2008**, *34* (4), 533-546. DOI: 10.1016/j.oregeorev.2008.03.004.
16. Couto, H.; Roger, G.; Moëlo, Y.; Bril, H. Le district à antimoine-or Dúrico-Beirão (Portugal): évolution paragenétique et géochimique; implications métallogéniques. *Mineralium Deposita* **1990**, *25* (1), S69-S81. DOI: 10.1007/BF00205252.
17. Couto, M. H. M. As mineralizações de Sb-Au da região Dúrico-Beirã. PhD thesis, Universidade do Porto, Porto, Portugal, **1993**.
18. Lotze, F. Zur Gliederung der Varisziden der Iberischen Meseta. *Geotekt. Forschg.* **1945**, *6*, 78-92.
19. Julivert, M.; Fontboté, J.; Ribeiro, A.; Conde, L. Mapa tectónico de la Península Ibérica, Canarias y Baleares, escala 1: 1.000. 000. IGME, Madrid, Spain: 1972.
20. Carvalho, A. Minas de Antimónio e Ouro de Gondomar. Estudos, Notas e Trabalhos do Serviço de Fomento Mineiro (1969) **1969**, XIX(1-2), 91-170.
21. Frutuoso, R. Soil Sampling Campaign Report Ribeiro da Serra Mine; [Unpublished Report] 2018.
22. Schwartz, G.; Eshel, G.; Ben Dor, E. Reflectance spectroscopy as a tool for monitoring contaminated soils. *Soil Contam* **2011**, *6790*. DOI: 10.5772/23661.
23. Dramsch, J.S. Chapter One - 70 years of machine learning in geoscience in review. In *Advances in Geophysics*, Moseley, B.; Krischer, L., Eds. Elsevier: 2020; Vol. 61, pp 1-55.
24. Ayodele, T. O. Machine learning overview. *New Advances in Machine Learning* **2010**, *2*, 9-18. DOI: 10.5772/9374
25. Cardoso-Fernandes, J.; Teodoro, A. C.; Lima, A.; Roda-Robles, E. Semi-automatization of support vector machines to map lithium (Li) bearing pegmatites. *Remote Sensing* **2020**, *12* (14), 2319. DOI: 10.3390/rs12142319.
26. Santos, D.; Cardoso-Fernandes, J.; Lima, A.; Müller, A.; Brönnner, M.; Teodoro, A. C. Spectral analysis to improve inputs to random forest and other boosted ensemble tree-based algorithms for detecting NYF pegmatites in Tysfjord, Norway. *Remote Sensing* **2022**, *14* (15), 3532. DOI: 10.3390/rs14153532.
27. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H. A.; Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering* **2018**, *31* (8), 1544-1554. DOI: 10.1109/TKDE.2018.2861006.
28. Cardoso-Fernandes, J.; Silva, J.; Dias, F.; Lima, A.; Teodoro, A. C.; Barrès, O.; Cauzid, J.; Perrotta, M.; Roda-Robles, E.; Ribeiro, M. A. Tools for Remote Exploration: A Lithium (Li) Dedicated Spectral Library of the Fregeneda-Almendra Aplite-Pegmatite Field. *Data* **2021**, *6* (3), 33. DOI: doi:10.3390/data6030033.
29. Zhou, W.; Yang, H.; Xie, L.; Li, H.; Huang, L.; Zhao, Y.; Yue, T. Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model. *Catena* **2021**, *202*. DOI: 10.1016/j.catena.2021.105222.

30. Baíllo, A.; Chacón, J.E. Chapter 1 - Statistical outline of animal home ranges: An application of set estimation. In Handbook of Statistics, Srinivasa Rao, A. S. R.; Rao, C. R., Eds. Elsevier: 2021; Vol. 44, pp 3-37.
31. Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* **2017**.
32. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022; p 568.
33. Wu, Y.; Chen, J.; Ji, J.; Gong, P.; Liao, Q.; Tian, Q.; Ma, H. A Mechanism Study of Reflectance Spectroscopy for Investigating Heavy Metals in Soils. *Soil Science Society of America Journal* **2007**, 71 (3), 918-926. DOI: 10.2136/sssaj2006.0285.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.