




Article

Convolutional Neural Networks Applied to Antimony Quantification via Soil Laboratory Reflectance Spectroscopy in 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 explored the feasibility of identifying and quantifying Sb mineralisations through the spectral signature of soils using laboratory reflectance spectroscopy, a non-invasive remote sensing technique, and by employing convolutional neural networks (CNNs). Standard signal pre-processing techniques were applied to the spectral data, and the soils were analysed by inductively coupled plasma mass spectrometry (ICP-MS). Despite achieving high R-squared (0.7) values and an RMSE of 173 ppm for Sb, the study faces a significant challenge of generalisation of the model to new data. Despite the limitations, this study provides valuable insights into potential strategies for future research in this field.



Citation: Carvalho, M.; Cardoso-Fernandes, J.; Lima, A.; Teodoro, A.C. Convolutional Neural Networks Applied to Antimony Quantification via Soil Laboratory Reflectance Spectroscopy in Northern Portugal: Opportunities and Challenges. *Remote Sens.* **2024**, *16*, 1964. <https://doi.org/10.3390/rs16111964>

Academic Editor: Eyal Ben-Dor

Received: 22 February 2024

Revised: 24 May 2024

Accepted: 28 May 2024

Published: 30 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: antimony; exploration; reflectance spectroscopy; soil analysis; critical raw materials; CNN; MobileNetv2; geology

1. Introduction

Antimony (Sb) is currently considered a critical raw material (CRM) to the European Union (EU), as it is strategic to its economy in a scenario where China dominates the global market of Sb. 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 sulphides are reported to have been used by humans since at least 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 A.D.), while “antimonium” is reported to be referred to by an Arabian alchemist, Geber, living in the eighth century [1,2]. 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 primary 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 indirectly detect heavy metals in soils using laboratory-based reflectance spectroscopy. Kemper and Sommer [7] used multiple linear regression (MLR) and artificial neural network (ANN) on reflectance spectra spanning 350–2500 nm to predict As, Cd, Cu, Fe, Hg, Pb, S, Sb, and Zn concentrations from samples collected after a mine tailings dam break in Spain. This accident resulted in a very contrasting mineralogy between the background soil and the contaminated zone. They report good predictions, with R^2 of 0.84, 0.72, 0.96, 0.95, 0.87, and 0.93 for As, Fe, Hg, Pb, S, and Sb, respectively, and of 0.51 for Cd, 0.43 for Cu, and 0.24 for Zn. Nanni and Dematté [8] utilised multiple regression equations on reflectance spectra (400–2500 nm) measured with a laboratory sensor to predict soil properties and obtained an R^2 above 0.79 for Fe_2O_3 and TiO_2 . Cheng, et al. [9] employed partial least squares regression (PLSR) on soil reflectance spectra (350–2500 nm) from a suburban area of Wuhan City, China, determining metal concentrations (Cd, Pb, As, Cr, Cu, and Zn) using inductively coupled plasma atomic emission spectroscopy (ICP-AES). Different pre-processing methods affected model performance, with the Savitzky–Golay treatment showing promise. Statistical analysis focused on relationships between soil reflectance spectra and SOM, Fe, and heavy metals. Results regarding internal relationships between heavy metal concentrations and spectrally active elements were inconclusive, warranting further study.

Rodríguez-Pérez, et al. [10] used PLSR on reflectance spectra (350–2500 nm) of air-dried soil samples to estimate Mn, soil nutrients, pH, and electrical conductivity from vineyards in Spain. They reported good performance only for phosphorus, pH, and electrical conductivity, with R^2 of 0.92 and above. Pyo, et al. [11] used a CNN and compared it to ANN and random forest regression (RFR) employed on reflectance spectra (749–2400 nm) to estimate As, Cu, and Pb from soil samples taken from a mining area located in the Geum River watershed of South Korea, and obtained the best results (R^2 of 0.86 for As, 0.74 for Cu, and 0.82 for Pb) with the CNN model, but also achieved reasonable soil heavy metal estimation accuracy with the other machine learning models. RFR was used by Guo, et al. [12] on reflectance spectra (350–2500 nm) to 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 content, obtaining the best results (R^2 of 0.60 for Zn, and 0.30 for Ni), 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 mineralisation through the spectral signature of soils. Previous studies have estimated heavy metals from soil laboratory spectroscopy, but Sb is not usually included in those studies. The exception is the work of [7], which predicted Sb taking into account the band centre and the full-width half-maximum related to Sb. However, the research methodology could not be reproduced, since specific band centres of the absorption features were not provided. This work addresses the attempt to use soil reflectance spectroscopy through an alternative approach, with the application of state-of-the-art algorithms like CNN, a deep learning technique, to make this estimation. Therefore, this is the first study to apply CNN to estimate soil Sb content.

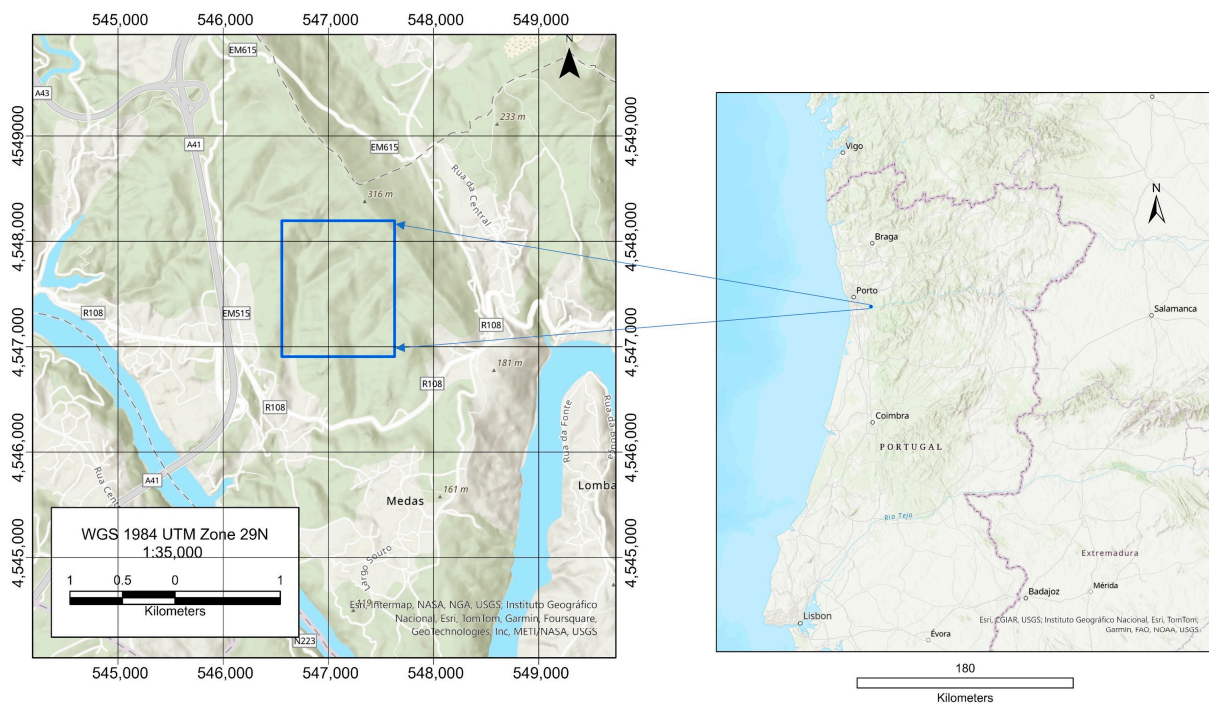
The major advantage of developing ML and deep learning (DL) methods to quantify and qualify heavy metals in soils is the possibility of analysing large-scale zones faster, avoiding the high cost and time demand that is implicit in the traditional approach of geochemical analysis. CNNs are being employed as nonlinear methods to capture features on the spectra to differentiate soil proprieties and offer advantages due to their capacity to identify patterns in data that humans cannot yet identify [15]. Padarian, et al. [16] applied deep learning to predict soil proprieties from a data set of 20,000 samples of raw spectral data from topsoils and obtained R^2 of 0.77 for total carbon, 0.72 for clay, and 0.58 for clay and sand, showing the potential for application in works on soil spectral data. Ng, et al. [17] compared the performance of PLSR, Cubist tree model, and CNN in predicting soil proprieties using spectral data from 14,569 samples. They had better performance with CNN mode, obtaining R^2 values above 0.9. However, they highlight that the application of these models can be limited due to the requirement of computational power, difficulty in preventing overfitting, and the requirement for large amounts of data.

Reflectance spectroscopy is a non-invasive remote sensing technique that is capable of identifying targets for mineral exploration, thus reducing costs and avoiding environmental impacts. In this study, reflectance spectroscopy was applied in the laboratory, although it can be employed in situ. Despite Sb itself not having specific absorption features in the visible and near-infrared to short-wave infrared (VNIR–SWIR) region, however, other soil proprieties present in the soil can serve as diagnostic features [13]. Sulphur has an absorption feature due to an electronic process called conduction bands [18,19]. Antimony and sulphur form a mineral called stibnite or antimonite, which occurs in the study area and can be present in its original form in the soils. Hunt [18] analysed the spectra of antimonite and described a well-defined conduction band in VNIR (300–800 nm). The advantage of CNN algorithms, like the one employed in this study, is that they can exploit feature representations that are learned exclusively from data and do not require hand-crafted features based on a priori knowledge [20]. Thus, the methodology employed relies on the principle that if antimony can be predicted by any feature(s) in the soils' reflectance spectra, the CNN can extract that information by itself by training on distinct samples with distinct antimony contents. Additionally, this study conducts a multivariate approach in which different elements analysed (As and Pb, for example) are used as pathfinders for Sb exploration.

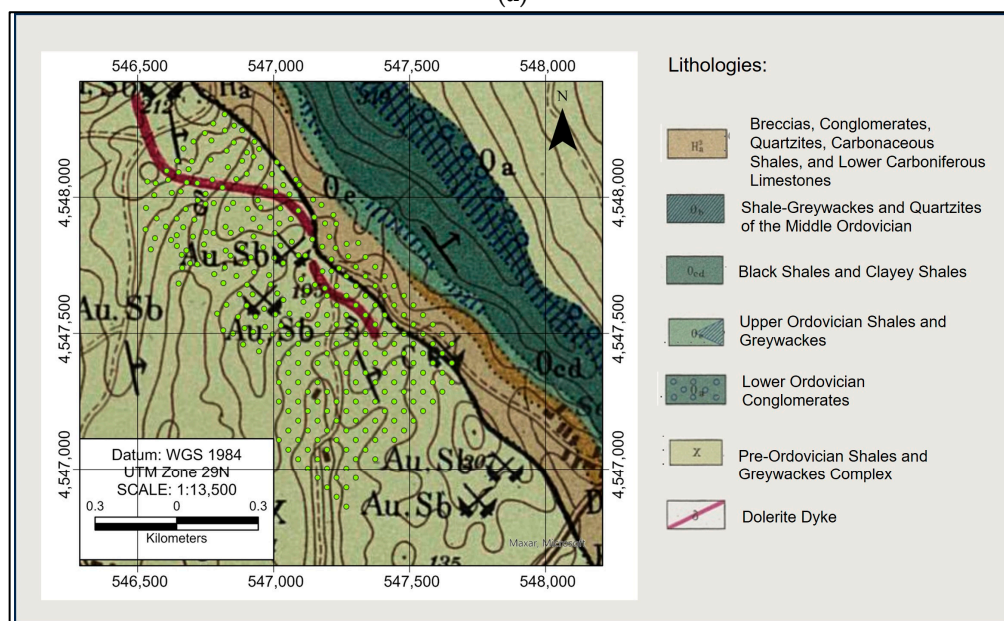
2. Background

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 of 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 [21,22]. 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 have been executed [23]. The mining structures are now abandoned (Figure 2), and many waste piles and tailings remain in the zone.



(a)



(b)

Figure 1. (a) Location of the study area; (b) 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 [22,24–26]. 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.



Figure 2. Ribeiro da Serra mine infrastructure pictured in the late 19th century [27] (a), and recent photo of ruins (b).

The Sb mineralisation occurs in low volume in discontinuous quartz veins that are mainly hosted in Silurian schists and greywackes [26]. 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 [21]. The dominant directions of the country rocks range from N to NW dipping to W. The most productive veins occur in the E–W direction dipping to N (Tapada) and N–S dipping to W (Ribeiro da Serra).

2.2. Spectral Reflectance

Reflectance spectroscopy offers a means to extract multiple soil properties, both direct and indirect, as well as metal content [28]. 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 VNIR–SWIR region, spanning from 400 to 2500 nm. This range is further divided into two sub-ranges: VNIR (400 to 1100 nm) and SWIR (1100 to 2500 nm). The interactions between light and matter are intricately tied to the wavelength. Though pure metals do not 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 [28]. Even if Sb does not have a specific feature that can be identified, this principle can be applied to estimate the Sb content through general parameters of the soil.

2.3. Convolutional Neural Network

The geosciences field has 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 [29–32] that are useful in geosciences. 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 [33]. The heterogeneity of the data and the multi-resolution are often associated with the already challenging multiconnected nature of geoscience processes [33]. Despite those challenges, the capacity for ML and DL methods to actuate in geoscience 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.

Currently, CNNs are the most used type of deep learning network [34–37]. The capacity of a CNN to capture nonlinear behaviours makes it suitable for geological problems. CNNs demonstrate 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 geoscience applications such as global water storage modelling, landslide prediction, and earthquake arrival time estimation [14,29]. The constitution of CNNs primarily involves three types of layers: convolutional layers, pooling layers, and fully connected layers [13].

3. Materials and Methods

3.1. Soil Collection and Preparation

The present study benefited from the soil samples collected in the scope of the AU-REOLE project-ERA-MIN/0005/2018 (<https://aureole.brgm.fr/>; accessed on 25 January 2024). The soil sampling campaigns covered 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 mineralisation distribution and identify possible new mineralised structures, while assessing soil contamination distribution. The first soil sampling campaign was carried out in the Ribeiro da Serra Sb-Au mine zone in 2021, while the second soil sampling campaign in the Tapada mine zone took place in 2022. In total, 309 samples were collected. The sampling campaigns follow a plan on a grid of 50×50 m in the area with an orientation according to the mining structures. Soils were collected from horizons B and C when there was no horizon B, which is relatively common in the study area. All samples underwent spectral signature collection and XRF analysis. From the totality of samples, 54 samples from Ribeiro da Serra and 53 samples from Tapada were sent to the Bureau Veritas laboratory in Vancouver (Canada), certified to ISO 17025 [38], to be analysed by ICP-MS (inductively coupled plasma mass spectrometry analysis). The data quality was assessed by inserting reference materials (STD OREAS45H, STD OREAS501D, and STD OREAS25A-4A), replicates, and blanks into randomly assigned positions within each analytical rack. Only a subset of the samples was chosen for laboratory analysis due to the elevated cost of such analysis. The selection of samples was performed randomly to ensure a representative subset. Before sending the samples to ICP-MS analysis, the samples were dried in a muffle furnace for at least 48 h and at a temperature of 55 °C. With the samples dry, after rifling, the soil samples were ground to a size $<200 \mu\text{m}$. In the study, we utilised only the samples sent to a certified laboratory for calibration purposes. From the samples sent to ICP-MS analysis, eight samples were sent in their totality to perform complementary analyses and were no longer available for reflectance spectroscopy studies. In the end, 99 samples were available to train and evaluate the model. After outlier removal (see Section 3.3), 92 samples were used in the present analysis. The samples with ICP-MS results for Sb above 1000 ppm were considered as outliers.

3.2. Soil Reflectance Measuring

The raw soil, previously dried, with gravel and pieces of plants removed, was used to take 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 five measurements was used, and each measurement resulted in 40 scans. A periodic wavelength check was performed using an external reference material (Mylar) to ensure instrument calibration. The standard deviation for the average

computation of the five measurements was assessed from the Mylar reference material, with a standard deviation ranging from 0.0001 to 0.0039, depending on the wavelength.

The FieldSpec 4 standard resolution spectroradiometer equipment (ASD Inc., Boulder, CO, USA) using a Contact Probe with an internal light source provided by a halogen bulb and a spot size of 10 mm was used to collect the spectral data in the laboratory. The spectroradiometer has three sensors (one VNIR and two SWIR sensors), with the spectral resolution of 3 nm at 700 nm, 10 nm at 1500, and 10 nm at 2100 nm [39]. All measurements were conducted in a dark room. The proceeding included “heating” the equipment for 30 min before starting the measurements as recommended by the manufacturer due to sensor sensitivity to ambient temperature. Heating time is necessary to ensure the three sensors are at the same temperature [39]. Moreover, splice correction was applied to smooth the splicing points between each instrument, resulting from the differences in temperature sensitivity of the three sensors. Normalisation 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 [39].

Selected spectra are graphically depicted in Figure 3, representing samples with varying Sb content. There is no correlation between the Sb content and the reflectance magnitude. However, there are visible differences between the spectral behaviour of the curves. There are no shoulders in the visible range for high Sb samples, and sample RS052 shows a ramp-like Fe^{2+} absorption feature in the VNIR [18]. There is also a tendency for samples with higher Sb content to show less pronounced absorption features at ~1910 nm and ~2200 nm, which correspond to the water and Al-OH features, respectively [40].

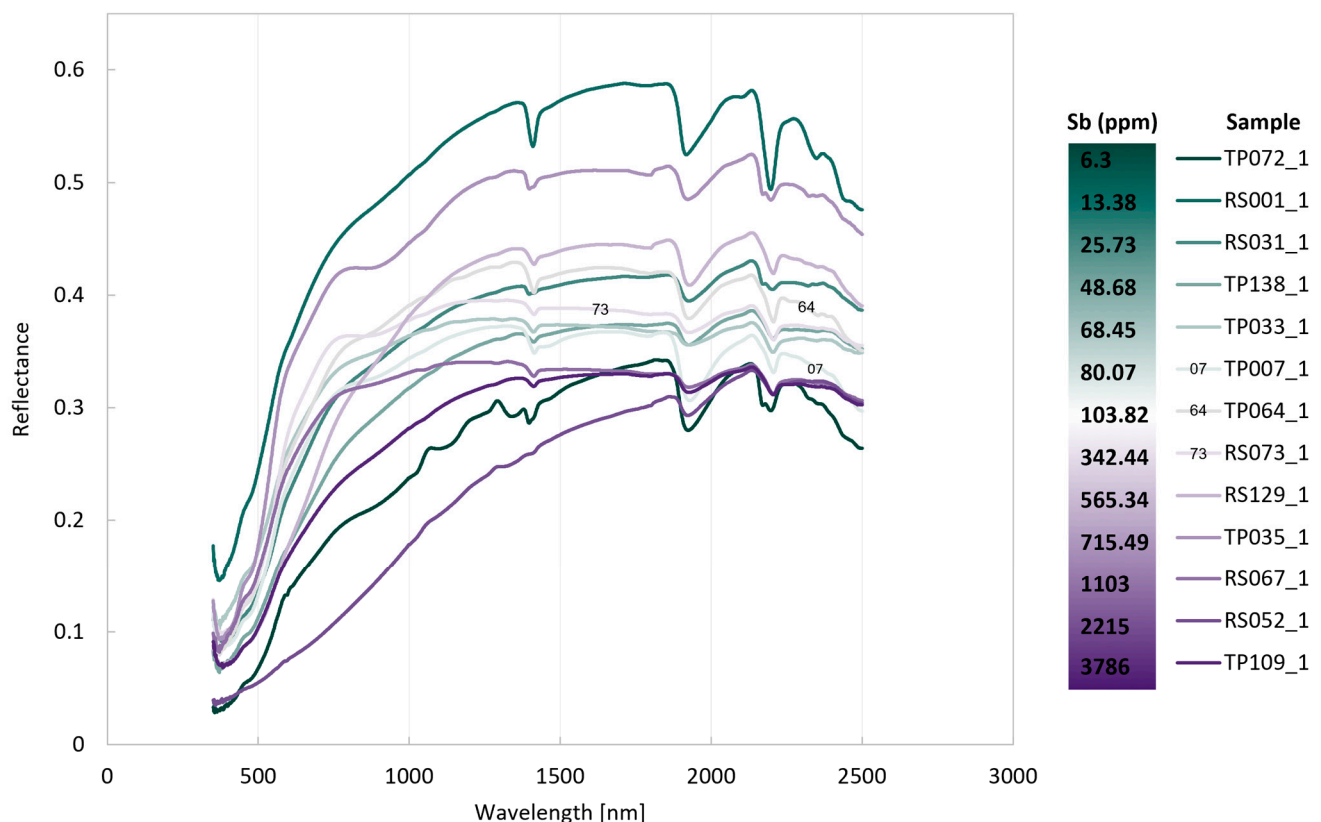


Figure 3. Examples of some recorded spectral data.

3.3. Spectral Pre-Processing

The first step was removing from the dataset samples with Sb content above 1000 ppm, obtained by ICP-MS analysis, which were considered outliers ($n = 7$) for being primarily associated with contaminated areas that do not represent the Sb concentrations found related to the natural occurrence of mineralisation (see Section 4.1), common spectral pre-processing 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: converting the reflectance to absorption; removing the continuum; smoothing the signal and calculating the first or second derivatives; and converting the waveform spectrogram.

The order of the spectral pre-processing is given in the diagram (Figure 4).

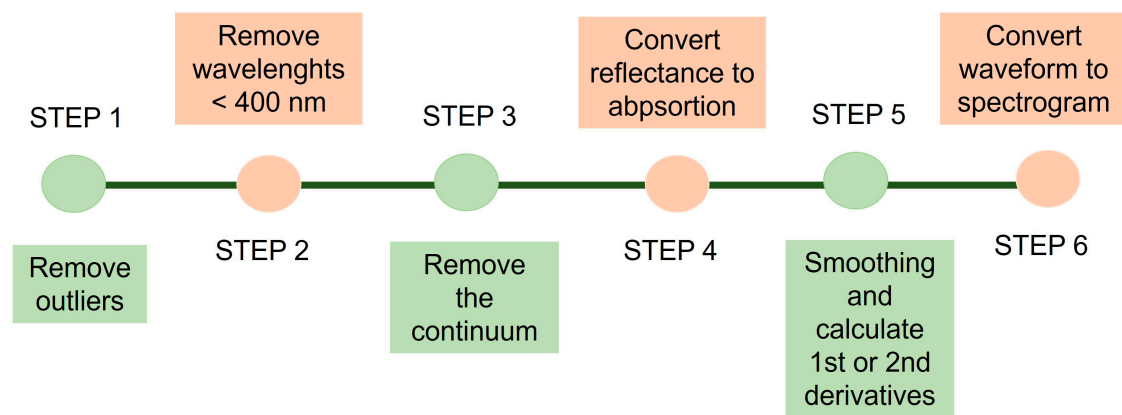


Figure 4. Pre-processing steps followed in this study.

3.3.1. Continuum Removal for Normalisation

Continuum removal is a standard technique that allows the extraction of characteristic absorption bands on the reflectance spectrum curves and the correct identification of the wavelength position of the absorption feature by eliminating the noise [41]. The convex hull forms a polygon connecting the outermost points within the sample while ensuring that this polygon's internal angles are less than 180 degrees, creating the smallest convex shape that encloses all the points in a given set [42]. The continuum-removed data were obtained using a Python script by [39].

3.3.2. Spectral Pre-Treatment

After the continuum removal, the absorption was calculated from the data using a Python function, as mentioned in [7], based on reflectance using the logarithmic relationship. After converting to absorption, the data were 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 are used to calculate the first or second derivatives by applying the `np.gradient` function from the NumPy library (Figure 5). This study set the polynomial degree to 2, and the window size was 31 nm.

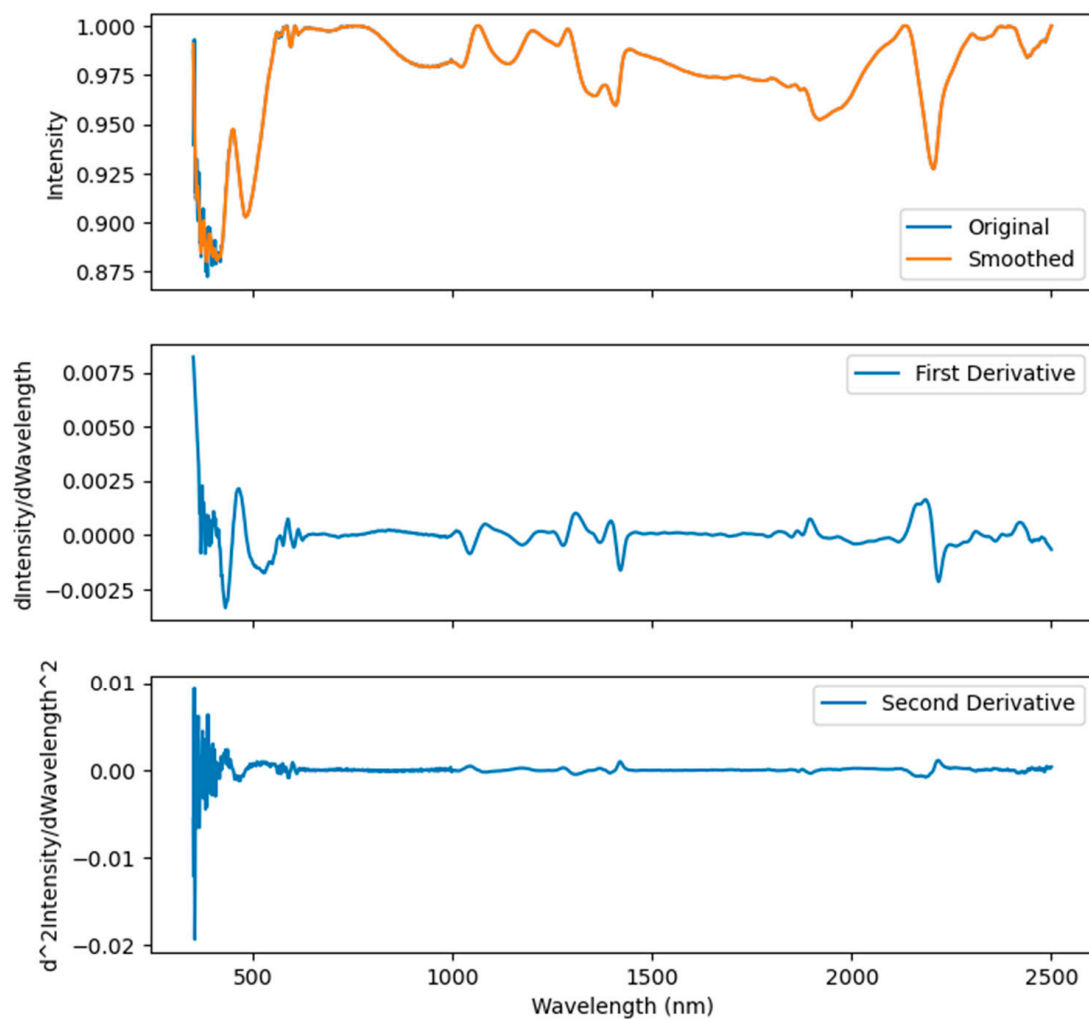


Figure 5. Spectral data and results of spectral pre-treatment before removing wavelengths between 350 and 400 nm (the noise is visible in the region).

3.3.3. Convert Waveform to Spectrogram

This step involves applying the short-time Fourier transform (STFT) to the `equal_length` tensor using `tf.signal.stft`. A new dimension is added to the spectrogram (initially 2D) tensor using `spectrogram[..., tf.newaxis]` (Figure 6). This dimension is included in 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 into three-dimensional data, as such structure is required for the application of a CNN. This methodology can be replicated by applying the Python script available in the Supplementary Materials (Code S1).

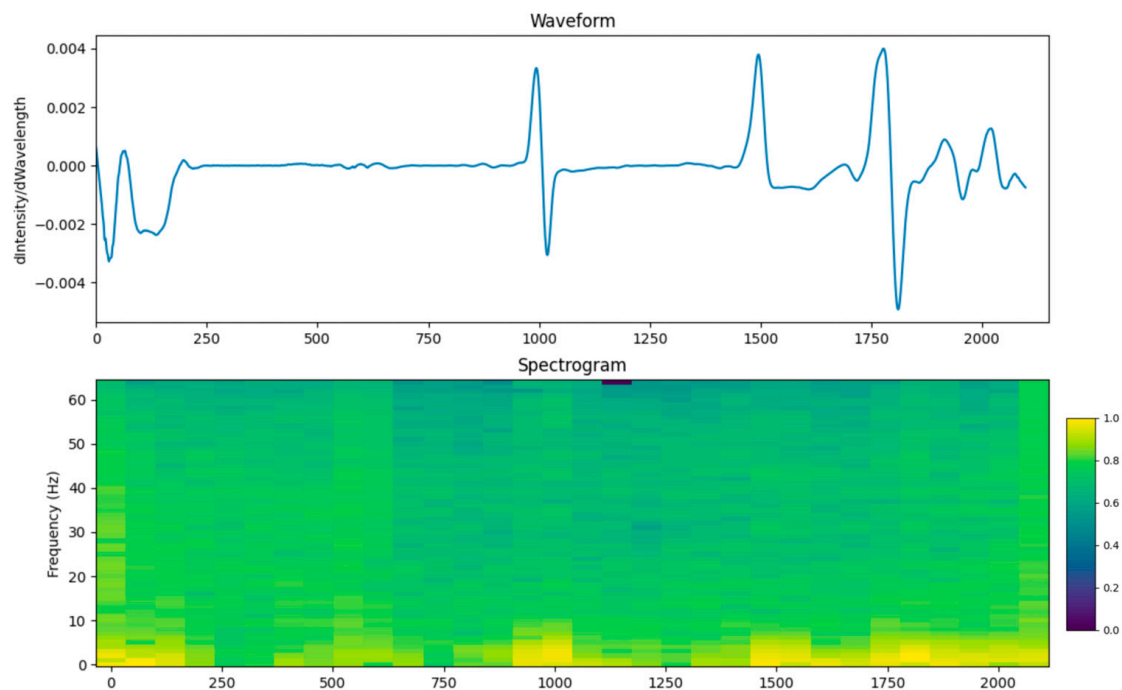


Figure 6. Transformation from waveform to spectrogram.

3.4. Application of Convolutional Neural Network

In the present work, to deal with limited computational capacity, a MobileNetV2 model was used. In a preliminary evaluation, PLSR and RF models were tested and compared with the CNN, which outperformed those methodologies [43], and in this study, only the results for this network are shown. 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 [34]. The model was chosen to be suitable for development on devices with limited resources, making it easier to apply in projects without requiring more sophisticated computational resources [44–46]. 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, and Zn, that also are not directly detected in the VNIR–SWIR spectral range and have exhibited different Person’s correlation with Sb in the study area (Table 1). The model input data consisted of the spectral data after conversion to spectrogram (see Code S1 and Supplementary Materials for additional information) and the target features corresponding to the selected element values. The function of the Sklearn library, `train_test_split` was employed to randomly split the data into two-thirds of the samples for training and one-third for validation, with a random state of 42.

Table 1. 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

The model's output data comprised an estimation of the element content for a given sample based on the provided spectrogram. During the model training, the batch size parameter was set to 64, limiting the number of input features, and the shuffle buffer size was set to 100. To diminish the overfitting, an early stopping mechanism was implemented. The MobileNetV2 [47] implemented corresponds to one convolution layer and three inverted residual blocks. Each convolutional block applies a 2D convolution operation followed by batch normalisation and ReLU6 activation (Figure 7).

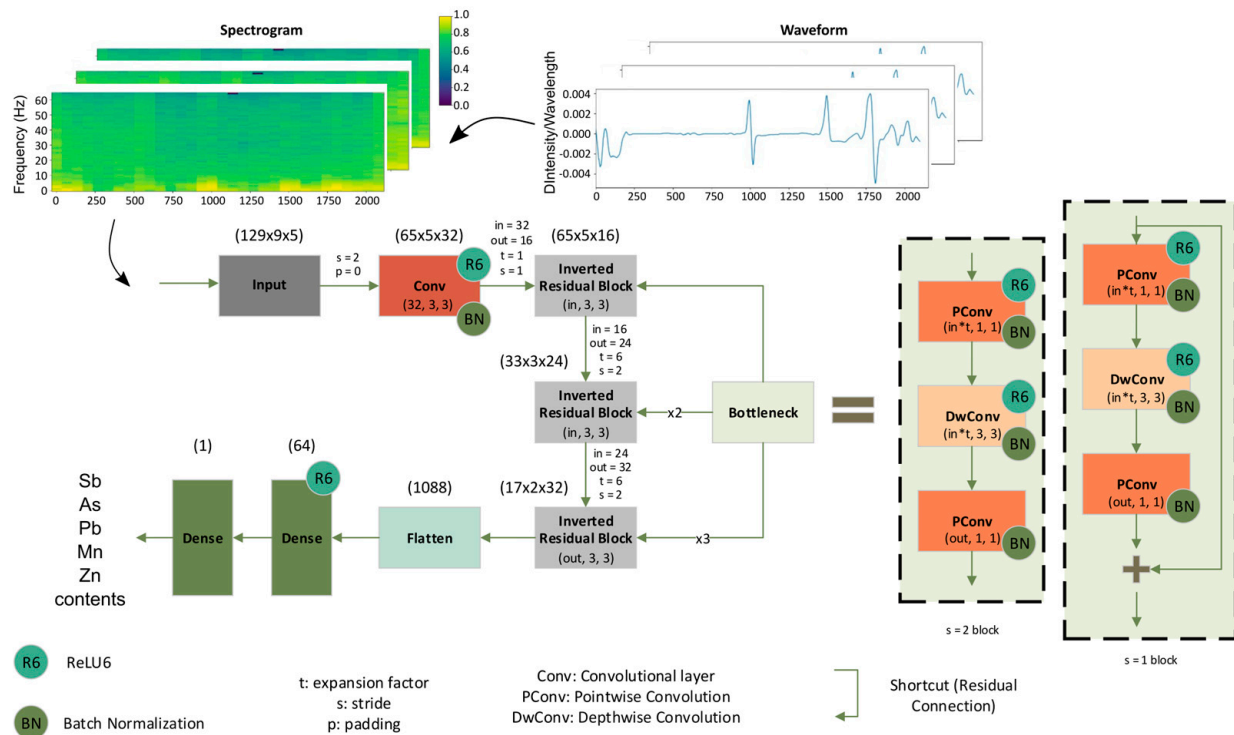


Figure 7. CNN structure and parameters employed in this study.

The metrics for evaluating the model performance were the R^2 and RMSE. Both are standard metrics for model performance evaluation [48]. R^2 , or coefficient of determination, is a measure that quantifies how much the independent variables explain the variance in the dependent variable in a regression model. It ranges from 0 to 1, where 0 indicates that the independent variables do not explain the variability in the dependent variable, and 1 indicates the independent variables can explain all the variability. R^2 is calculated as the ratio of the explained variance to the total variance (Equation (1)) [49].

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (1)$$

RMSE is a natural derivation of the mean squared error, which calculates the averaged squared difference between the predicted values and the actual values [49] and can be interpreted in the same units as the original data, helping to access the typical magnitude of errors made by the model. RSME is given by Equation (2).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (2)$$

4. Results

4.1. ICP-MS Analysis

The results obtained from the ICP-MS analysis are shown in Table A1, which indicates which samples were discarded from the training set. The distribution of Sb concentrations in the study area is depicted in Figure 8. The descriptive statistics for the selected elements are also presented (Table 2).

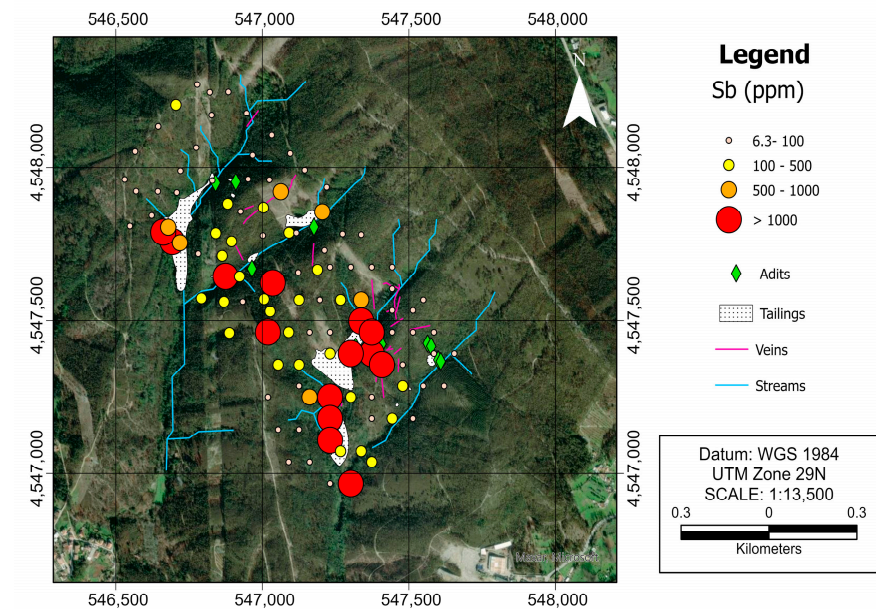


Figure 8. The distribution of Sb concentrations in the study area, the position of old mining adits, the known veins, tailings, and local streams.

Table 2. Descriptive statistics for the selected elements.

	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)
Mean	132	69	30	75	36
Std. Deviation	184	156	38	123	23
Minimum	6.3	15	9	9	11
Maximum	895	1431	360	844	146
Q1	26	23	20	27	23
Q2	59	34	24	39	30
Q3	155	69	29	72	42

Note: Q1, Q2, Q3 refer to first, second, and third quartiles.

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, but 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 are close 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.

4.2. Spectral Pre-Processing and CNN Results

Removing the outliers based on the Sb concentrations and training the model with only the samples that contained up to 1000 ppm of Sb content had a positive impact on the

performance of the model, leading to higher values for R^2 when the values with the outliers were 0.4 for R^2 and 700 ppm for RMSE. Similarly, the application of the pre-processing 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 pre-processing 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 pre-processing methods are presented. These results correspond to the signal used as input (wavelengths between 400 and 2400 nm), using the reflectance converted to absorption; spectra after removing the continuum; and signal smoothing and calculation of the first derivative. The input was this processed signal converted to a waveform spectrogram. Training the model using multiple elements instead of a single element was tested for making the predictions. Still, 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 R^2 , RMSE (ppm) for train and validation and the number of training epochs.

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 overfitting issue. The model learned to predict the values of Sb for the training set but with a high validation error (Figure 9).

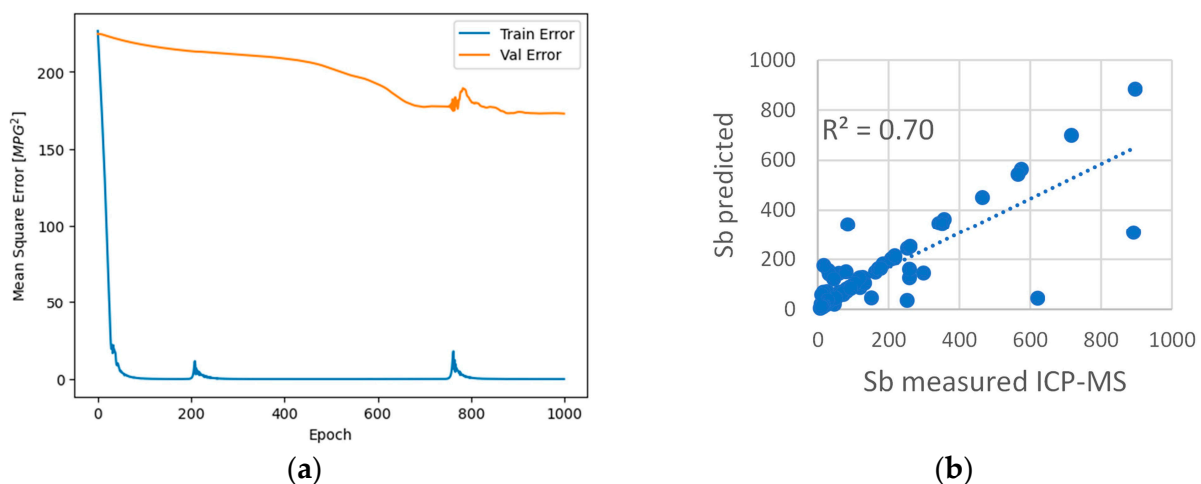


Figure 9. (a) Train error versus validation error by epoch for Sb. (b) R^2 (train and test set) for Sb predicted and measured.

Overfitting occurs when the model learns the training data too well, capturing patterns specific to the training set but failing to generalise well to new, unseen data [50]. 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 10).

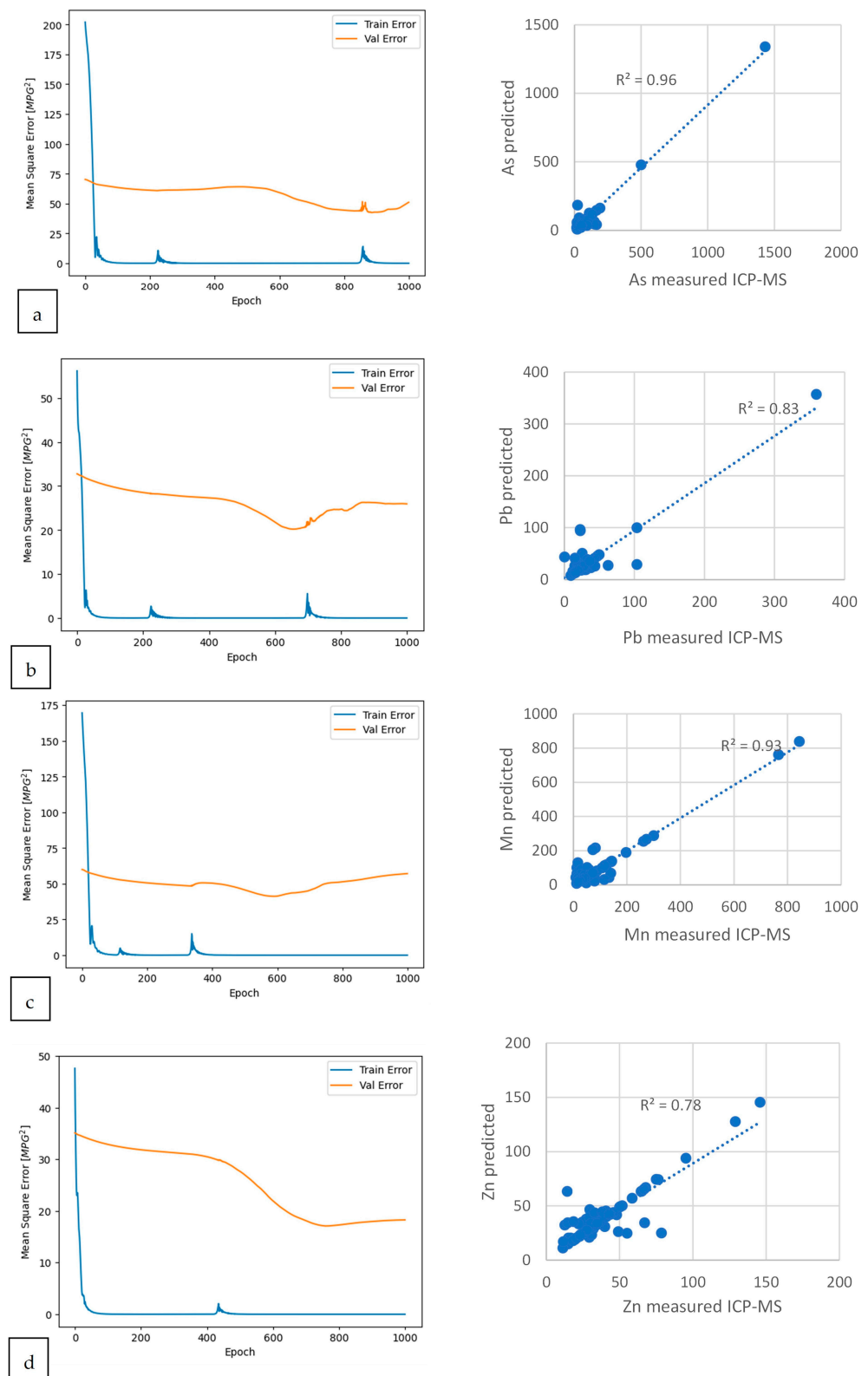


Figure 10. (Left) Train error versus validation error by epoch and (right) R^2 for train and test set for (a) As, (b) Pb, (c) Mn, and (d) Zn.

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 could not achieve a good generalisation performance. However,

some elements experiment better results for the generalisation, namely As and Mn. The overfitting tendency is more pronounced for Sb and Zn. We can observe in the graphics in Figure 10 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 at a plateau. It is worth mentioning that while the validation error may appear to have a tendency to decrease, tests executed with a few thousand epochs more show that there is no improvement in the predictions. Those results may indicate that there is a limit in the generalisation that the model can reach with the present data set.

5. Discussion

The soil samples for this study were obtained and analysed in previous studies to characterise the Sb distribution in the former mining areas of Ribeiro da Serra and Tapada, in Northern Portugal. The existence of high-concentration samples (outliers) is an inherent issue of soil spectroscopy, although they make regression problems challenging to solve. Removal of outliers is a standard practice in chemometric studies based on reflectance spectroscopy [51,52]. In our research, outliers represented only seven samples, but removing them from the dataset improved R^2 for the Sb prediction from 0.4 to 0.7 and the RMSE from 700 ppm to 173 ppm. The continual removal, smoothing, and application of the first derivative also improved the results, but not the application of the second derivative.

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 improve 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] used a methodology to degrade the spectra considering the band centre and the full-width half-maximum to resolve the overfitting issue. The band centre 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 centre for the target element is unknown. In addition, in their study area, they have a big contrast between the region's soil 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 those that do not; the Sb content is relatively discrete. Wu, et al. [53] found that the correlation with total Fe, active and residual, was a major predictive mechanism for heavy metals in soils. OM and clay also have a correlation. The soil analysis in the current study did not include those proprieties, which can be a way to obtain better results.

As the soil sampling campaigns executed focused on capturing the general distribution of the Sb, they did not capture the progressive increment in Sb in mineralised 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 the progressive content of Sb in the soils associated with the veins could be more appropriate for this study and work as a solution for the overfitting in the CNN model. In addition, obtaining OM and clay data from the soils can be a different approach that can help better understand the features in the spectral signature of soils containing Sb that can be employed for its identification.

Although this study applies a light CNN that can be easily implemented without special computational demands, there are limitations in the application due to the small number of samples available, as is expected in this kind of methodology, a high number of samples is often necessary. Nonetheless, the sample size used in this study aligns with what is mentioned in the literature [7,54–59]. Some studies even present smaller datasets [51,60,61]. Another limitation is that there are no specific features in the literature to identify the absorption of Sb in soils. Also, like all neural networks, the CNN is a

black-box method. Even though CNN can learn from patterns in the data, and despite its application in a Python environment allowing greater control of the model's parameters, the specific wavelengths that contribute the most to the model's prediction cannot be directly accessed, although [62] did determine active areas of the spectra through the activation filters. Future studies can address the poor interpretability of neural network models by using post-hoc techniques, such as SHapley Additive exPlanations (SHAP), based on game theory, that explain the output of machine learning models and allow visualisation of the features in which the convolutional layers focus [62]. Other methods like RF, while still being a black-box method, allow the assessment of feature importance on the model's performance. On the other hand, PLSR is not a black box, with transparency on how the prediction is achieved [61]. Both RF and PLSR were previously accessed, but they did not achieve satisfactory results and were outperformed by CNN [43].

Despite having a larger number of samples available, not all samples were analysed in the laboratory due to the high costs of such analysis, which adds to the study's limitations. It is advised that future research should incorporate more samples. The CNN model shows promising results, but in this study, the overfitting of the model could not be avoided. Despite not having highly satisfactory results for the Sb predictions, this study provides insights about which strategies could be incorporated into future studies. This study did not have access to parameters such as SOM, texture, organic matter, and clay from the soils. Future studies could incorporate those parameters to better understand the features in the spectral signature of soils containing Sb. Additionally, those parameters can be useful for exploring multi-input and multi-output CNNs [59] and enhancing the model robustness and simultaneous multiple pre-treatments. We previously explored [43] various combinations of input elements with multi-elements, but the prediction accuracy for the individual elements was degraded when this multiple input was applied.

6. Conclusions

This study found varying concentrations of Sb in the sampled area, with the higher values of Sb influenced by historical mining activities and potential contamination. Implementing a CNN with low computational demand, MobileNetV2 model, for predicting Sb values shows promising results with a good fit for the training data, but with issues in generalising to new data. However, challenges emerged regarding its ability to generalise to new data. Notably, pre-processing steps remain essential for enhancing model performance. Future studies should consider alternative sampling methodologies and the increment and diversification of the available dataset, as the incorporation of other soil properties such as OM and clay into the analysis could provide more insight into the topic. This study provides insights into applying CNNs to predict Sb concentrations using spectral data while challenges remain to overcome.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16111964/s1>, Code S1: waveform to spectrogram; Code S2: Convolutional neural network.

Author Contributions: Conceptualisation, 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., and A.L.; visualisation, 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 Materials. Spectra continuum removal and absorption extraction were accomplished using a Python routine pub-

licly available at <https://www.mdpi.com/2306-5729/6/3/33/s1> (accessed 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, Christian Therien, licensed under an Apache License Version 2.0. and available on GitHub repository <https://github.com/ctherien>, accessed on 19 February 2024). Spectral data used in this paper are 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, MDL (minimum detection limit) of the analysis for the given elements, BLK (blank reference), STD (Stardart) reference materials and pulp duplicates. The standard deviation for duplicate samples is below 6 ppm, and the relative standard deviation is below 3%.

Sample	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)	Sample	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)
MDL	0.02	0.2	0.02	1	0.2	MDL	0.02	0.2	0.02	1	0.2
RS001	13.38	25	33.49	13	12.7	TP004	298.55	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	184.61	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	357.04	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	715.49	78	49.92	142	68
RS031	25.73	20.3	20.93	57	29.8	TP038	116.33	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	259.18	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	207.65	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	103.82	34.8	33.38	106	65.3
RS055	50.68	21.2	12.64	15	23.6	TP065	352.47	71	22.64	26	26
RS057	30.04	20.8	20.98	73	39.7	TP068	217.94	47.9	20.33	37	31
RS061	82.11	30.8	25.48	24	25.9	TP070	891.09	99.7	25.65	132	67
RS066 *	>4000	501.9	197.01	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	258.22	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	171.89	143	51.8	TP084	22.03	23.9	32.51	82	40.8

Table A1. Cont.

Sample	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)	Sample	Sb (ppm)	As (ppm)	Pb (ppm)	Mn (ppm)	Zn (ppm)
MDL	0.02	0.2	0.02	1	0.2	MDL	0.02	0.2	0.02	1	0.2
RS084	59.11	21.1	17.59	38	22.3	TP087	619.71	92.3	103.39	78	38
RS088	89.74	34.8	20.79	30	22.8	TP091	464.24	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	1040.14	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	103.51	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	125.76	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	449.03	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	359.75	88	26.7	TP136	33.92	42	27.22	72	43.5
RS131 *	>4000	895.5	228.86	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	221.68	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	176.98	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	BLK	<0.02	0.3	0.04	<1	0.3
		Reference Material STD OREAS45H					1.12	16.3	11.39	416	38.9
		Reference Material STD OREAS501D					2.56	13.4	24.43	379	81.7
		Reference Material STD OREAS45H					0.84	16.1	11.49	426	39.3
		Reference Material STD OREAS501D					2.43	11.4	24.3	371	83.2
		Soil Pulp RS090					101.96	76.7	22.39	48	25
		Soil Replicate RS090					96.97	76.5	22.25	46	24.9
		Soil Pulp RS037					575.47	1431.2	47.29	76	49.8
		Soil Replicate RS037					564.07	1426.4	48.43	77	50.8

* Samples excluded from the training set.

References

- Li, T.; Archer, G.F.; Carapella, S.C., Jr. Antimony and Antimony Alloys. In *Kirk-Othmer Encyclopedia of Chemical Technology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2000; pp. 1–15.
- Butterman, W.; Hilliard, H. *Mineral Commodity Profiles. Selenium*; Rapport US Department of the Interior US Geological Survey: Online Only, 2004; pp. 1–20. [\[CrossRef\]](#)
- Wisniak, J. Nicolas Lémy. *Rev. CENIC Cienc. Químicas* **2005**, *36*, 123–130.
- 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: Luxembourg, 2023; Available online: <https://commission.europa.eu/about-european-commission/departments-and-executive-agencies/internal-market-industry-entrepreneurship-and-smes> (accessed on 27 May 2024).
- Moolayadukkam, S.; Bopaiah, K.A.; Parakkandy, P.K.; Thomas, S. Antimony (Sb)-Based Anodes for Lithium–Ion Batteries: Recent Advances. *Condens. Matter* **2022**, *7*, 27. [\[CrossRef\]](#)
- He, J.; Wei, Y.; Zhai, T.; Li, H. Antimony-based materials as promising anodes for rechargeable lithium-ion and sodium-ion batteries. *Mater. Chem. Front.* **2018**, *2*, 437–455. [\[CrossRef\]](#)
- Kemper, T.; Sommer, S. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* **2002**, *36*, 2742–2747. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nanni, M.R.; Demattè, J.A.M. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Sci. Soc. Am. J.* **2006**, *70*, 393–407. [\[CrossRef\]](#)

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. [\[CrossRef\]](#)
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*, 1895. [\[CrossRef\]](#)
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. [\[CrossRef\]](#) [\[PubMed\]](#)
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 Sens.* **2022**, *14*, 5804. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
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. *Artif. Intell. Earth Syst.* **2022**, *1*, e220012. [\[CrossRef\]](#)
15. Wang, Y.; Abliz, A.; Ma, H.; Liu, L.; Kurban, A.; Halik, Ü.; Pietikäinen, M.; Wang, W. Hyperspectral Estimation of Soil Copper Concentration Based on Improved TabNet Model in the Eastern Junggar Coalfield. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–20. [\[CrossRef\]](#)
16. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning to predict soil properties from regional spectral data. *Geoderma Reg.* **2019**, *16*, e00198. [\[CrossRef\]](#)
17. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* **2019**, *352*, 251–267. [\[CrossRef\]](#)
18. Hunt, G.R. Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics* **1977**, *42*, 501–513. [\[CrossRef\]](#)
19. Clark, R.N. Spectroscopy of rocks and minerals and principles of spectroscopy: Chapter 1. In *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3rd ed.; Ryerson, R.A., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1999.
20. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1264. [\[CrossRef\]](#)
21. Neiva, A.M.R.; Andráš, P.; Ramos, J.M.F. Antimony quartz and antimony–gold quartz veins from northern Portugal. *Ore Geol. Rev.* **2008**, *34*, 533–546. [\[CrossRef\]](#)
22. Couto, H.; Roger, G.; Moëlo, Y.; Bril, H. Le district à antimoine-or Dúrico-Beirão (Portugal): Évolution paragenétique et géochimique; implications métallogéniques. *Miner. Depos.* **1990**, *25*, S69–S81. [\[CrossRef\]](#)
23. Couto, M.H.M. As mineralizações de Sb-Au da região Dúrico-Beirã. Ph.D. Thesis, Universidade do Porto, Porto, Portugal, 1993.
24. Lotze, F. Zur Gliederung der Varisziden der Iberischen Meseta. *Geotekt. Forschg.* **1945**, *6*, 78–92.
25. 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.
26. Carvalho, A. *Minas de Antimónio e Ouro de Gondomar. Estudos, Notas e Trabalhos do Serviço de Fomento Mineiro* (1969); Serviço de Fomento Mineiro: Lisboa, Portugal, 1969; Volume XIX, pp. 91–170.
27. Frutuoso, R. Soil Sampling Campaign Report Ribeiro da Serra Mine. 2018; [Unpublished Report of Aureole project (10.54499/ERA-MIN/0005/2018)].
28. Schwartz, G.; Eshel, G.; Ben Dor, E. Reflectance spectroscopy as a tool for monitoring contaminated soils. In *Soil Contamination*; IntechOpen Limited: London, UK, 2011; Volume 6790.
29. Dramsch, J.S. 70 years of machine learning in geoscience in review. *Adv. Geophys.* **2020**, *61*, 1–55.
30. Ayodele, T.O. Machine learning overview. *New Adv. Mach. Learn.* **2010**, *2*, 9–18. [\[CrossRef\]](#)
31. Cardoso-Fernandes, J.; Teodoro, A.C.; Lima, A.; Roda-Robles, E. Semi-automatization of support vector machines to map lithium (Li) bearing pegmatites. *Remote Sens.* **2020**, *12*, 2319. [\[CrossRef\]](#)
32. 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 Sens.* **2022**, *14*, 3532. [\[CrossRef\]](#)
33. Karpadne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 1544–1554. [\[CrossRef\]](#)
34. 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* **2017**, arXiv:1704.04861.
35. Kumar Lilhore, U.; Simaiya, S.; Sharma, Y.K.; Kaswan, K.S.; Rao, K.B.; Rao, V.M.; Baliyan, A.; Bijalwan, A.; Alroobaea, R. A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization. *Sci. Rep.* **2024**, *14*, 4299. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 4712. [\[CrossRef\]](#)
37. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [\[CrossRef\]](#) [\[PubMed\]](#)

38. ISO/IEC 17025:2005; General Requirements for the Competence of Testing and Calibration Laboratories. 3rd ed. International Organisation for Standardisation: Geneva, Switzerland, 2017.
39. 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*, 33. [\[CrossRef\]](#)
40. Pontual, S.; Merry, N.; Gamson, P. *G-Mex Spectral Interpretation Field Manual*; AusSpec International: Sydney, Australia, 1997.
41. 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*, 105222. [\[CrossRef\]](#)
42. Baíllo, A.; Chacón, J.E. Statistical outline of animal home ranges: An application of set estimation. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2021; Volume 44, pp. 3–37.
43. Carvalho, M. Machine Learning Applied to Sb Mineralizations in Northern Portugal. Master's Thesis, Faculdade de Ciências da Universidade do Porto, Porto, Portugal, 2023.
44. Rybczak, M.; Kozakiewicz, K. Deep Machine Learning of MobileNet, Efficient, and Inception Models. *Algorithms* **2024**, *17*, 96. [\[CrossRef\]](#)
45. Wang, H.; Qiu, S.; Ye, H.; Liao, X. A Plant Disease Classification Algorithm Based on Attention MobileNet V2. *Algorithms* **2023**, *16*, 442. [\[CrossRef\]](#)
46. Dokl, M.; Van Fan, Y.; Vujanović, A.; Pintarič, Z.N.; Aviso, K.B.; Tan, R.R.; Pahor, B.; Kravanja, Z.; Čuček, L. A waste separation system based on sensor technology and deep learning: A simple approach applied to a case study of plastic packaging waste. *J. Clean. Prod.* **2024**, *450*, 141762. [\[CrossRef\]](#)
47. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [\[CrossRef\]](#)
48. Hodson, T.O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [\[CrossRef\]](#)
49. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [\[CrossRef\]](#)
50. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022; p. 568.
51. Oliveira, D.L.B.; Pereira, L.H.d.S.; Schneider, M.P.; Silva, Y.J.A.B.; Nascimento, C.W.A.; van Straaten, P.; Silva, Y.J.A.B.; Gomes, A.d.A.; Veras, G. Bio-inspired algorithm for variable selection in i-PLSR to determine physical properties, thorium and rare earth elements in soils from Brazilian semiarid region. *Microchem. J.* **2021**, *160*, 105640. [\[CrossRef\]](#)
52. Kopačková-Strnadová, V.; Rapprich, V.; McLemore, V.; Pour, O.; Magna, T. Quantitative estimation of rare earth element abundances in compositionally distinct carbonatites: Implications for proximal remote-sensing prospecting of critical elements. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102423. [\[CrossRef\]](#)
53. 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 Sci. Soc. Am. J.* **2007**, *71*, 918–926. [\[CrossRef\]](#)
54. Xuemei, L.; Jianshe, L. Using short wave visible–near infrared reflectance spectroscopy to predict soil properties and content. *Spectrosc. Lett.* **2014**, *47*, 729–739. [\[CrossRef\]](#)
55. Gomez, C.; Viscarra Rossel, R.A.; McBratney, A.B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma* **2008**, *146*, 403–411. [\[CrossRef\]](#)
56. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [\[CrossRef\]](#)
57. Bajorski, P.; Kazmierowski, C.; Cierniewski, J.; Piekarczyk, J.; Kusnierek, K.; Królewicz, S.; Terelak, H.; Stuczynski, T.; Maliszewska-Kordybach, B. Use of clustering with partial least squares regression for predictions based on hyperspectral data. In Proceedings of the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014. [\[CrossRef\]](#)
58. Dunn, B.; Batten, G.; Beecher, H.G.; Ciavarella, S. The potential of near-infrared reflectance spectroscopy for soil analysis—A case study from the Riverine Plain of south-eastern Australia. *Aust. J. Exp. Agric.* **2002**, *42*, 607–614. [\[CrossRef\]](#)
59. Wang, C.; Zhang, T.; Pan, X. Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. *Geoderma* **2017**, *306*, 120–126. [\[CrossRef\]](#)
60. Jung, A.; Vohland, M. Snapshot Hyperspectral Imaging for Soil Diagnostics—Results of a Case Study in the Spectral Laboratory. *Photogramm.-Fernerkund.-Geoinf.* **2014**, 511–522. [\[CrossRef\]](#)
61. Henseler, J. Partial least squares path modeling: Quo vadis? *Qual. Quant.* **2018**, *52*, 1–8. [\[CrossRef\]](#)
62. Tsakiridis, N.L.; Keramaris, K.D.; Theocharis, J.B.; Zalidis, G.C. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma* **2020**, *367*, 114208. [\[CrossRef\]](#)

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.