

Article

Magnetite Talks: Testing Machine Learning Models to Untangle Ore Deposit Classification—A Case Study in the Ossa-Morena Zone (Portugal, SW Iberia)

Pedro Nogueira ^{1,2,*}  and Miguel Maia ²

¹ Department of Geosciences, School of Science and Technology, Évora University, Rua Romão Ramalho 59, 7000-761 Évora, Portugal

² Institute of Earth Sciences—Évora Pole, Rua Romão Ramalho 59, 7000-761 Évora, Portugal; mcmaiageo@gmail.com

* Correspondence: pmn@uevora.pt

Abstract: A comprehensive investigation into the application of machine learning algorithms for accurately classifying mineral deposit types is presented. The study specifically focuses on iron deposits in the Portuguese Ossa-Morena Zone, employing a limited dataset of trace element geochemistry from magnetites. The research aims to derive meaningful methodological and metallogenic conclusions from the obtained results. The findings demonstrate that the combination of a restricted dataset of trace element geochemistry from magnetites with diverse machine learning models serves as a reliable tool for achieving precise classifications of mineral deposit types. Among the machine learning methods evaluated, random forest, naïve Bayes, and multinomial logistic regression emerge as the most accurate classifiers, whereas the support vector machine, the k-nearest neighbour, and artificial neural networks exhibit lower performance scores. By integrating all literature-proposed classifications, and applying them to selected iron deposits, confident classifications were obtained. Alvito and Azenhas are reliably classified as skarns, whereas Monges, Serrinha, and Vale da Arca are classified as either porphyry or a Banded Iron Formation (BIF). Notably, the classification of Orada proves cryptic, encompassing both BIF and volcanogenic massive sulphide (VMS) deposit types. Moreover, the application of machine learning models to pertinent case studies offers valuable insights not only for classifying mineral deposit types but also for discerning mixed or complex origins. This approach provides meaningful results that can aid in the interpretation of mineral deposit types and may facilitate the identification of new mineral exploration targets. The research highlights the robustness of machine learning algorithms in interpreting magnetite data and underscores their potential significance in exploration projects.

Keywords: mineral deposit classification; trace element geochemistry; magnetites; LA-ICP-MS; Ossa-Morena Zone



Citation: Nogueira, P.; Maia, M. Magnetite Talks: Testing Machine Learning Models to Untangle Ore Deposit Classification—A Case Study in the Ossa-Morena Zone (Portugal, SW Iberia). *Minerals* **2023**, *13*, 1009. <https://doi.org/10.3390/min13081009>

Academic Editor: Paul Alexandre

Received: 28 June 2023

Revised: 25 July 2023

Accepted: 26 July 2023

Published: 29 July 2023



Copyright: © 2023 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/>).

1. Introduction

One challenging task in the exploration of ore deposits is defining the model that best fits a certain deposit type, based on the combination of multiple sources of data. The accurate definition of such models is a key factor for driving mineral exploration campaigns, helping to define the applicable methodologies and strategies, thus saving time and money.

Magnetite is a common mineral present in multiple metallogenic settings, so it is a convenient target to perform such a task. Its trace element composition provides valuable information to constrain the conditions and mechanisms associated with ore deposition, which can ultimately distinguish between barren and mineralized areas. With such goals in mind, magnetite trace element data have competently been used in the classification of ore deposits with significant developments in recent years (e.g., [1–6]). Nevertheless,

the chemical analysis of magnetite often ends in large datasets with numerous variables, constituting a challenge for the scrutiny of the results.

The determination of magnetite composition with laser-ablation–inductively-coupled-plasma-mass-spectrometry (LA-ICP-MS) has been proven to contribute to the correct classification of ore deposits using conventional discriminatory diagrams [7–10].

The classical approach [7] in the use of magnetite composition as a geochemical tool relies on a multi-phased analysis of the data, priorly separating some deposit types and applying discriminant diagrams that provide an affinity to a certain ore deposit type.

Data reduction methods, such as a Principal Component Analysis (PCA), have provided good results [5,6,11,12]. Recently, the application of machine learning (ML) supervised classification methods to the trace element composition of magnetite was used to classify ore deposits (e.g., [13–15]). However, these models are usually dependent on large datasets (e.g., ~17 k observations) to provide results [15].

The use of ML in the classification of ore deposits based on magnetite trace element data presents a huge advantage when compared to classifications based in conventional scatter plot diagrams. ML methods incorporate the multi-dimensional characteristics of the datasets, providing a more robust inspection of data and, consequently, a better aggregation of the different groups that can be used for defining the geological models.

For testing this assertion, the performance of six ML classification models is applied, and the results are discussed with a Reference Dataset with a much smaller number of data values (271 observations) of a magnetite trace element analysis published by different authors. Afterwards, the models are applied to a Case Study Dataset of magnetite composition from an area that comprises several iron deposits from the Portuguese sector of the Ossa-Morena Zone (OMZ), and the results are discussed based on the known geological models. The magnetite data gathered in the Case Study Dataset belong to three volcanogenic-exhalative deposits (i.e., Monges, Vale da Arca, Serrinha) from the Montemor-o-Novo Iron Complex (MIC), complemented by three Fe-skarn deposits (i.e., Alvito, Azenhas, and Orada). This study presents a novel approach in the classification of ore deposits in Portugal by applying six ML classification models to magnetite trace element analysis data. Notably, there is no record of such a methodology being previously employed in this specific context.

While the use of a limited dataset is acknowledged in this study, the results obtained from the machine learning classification of magnetites demonstrate promising levels of generalizability. Despite the dataset's limitations, the model's ability to compare favourably with conventional classification diagrams [7,9,10,14] suggests that it can effectively classify magnetite samples from various ore deposit types.

It is important to note that while the dataset's size may be limited, efforts were made to ensure its representativeness and diversity, capturing essential characteristics of magnetites from different origins and evolutions. Additionally, rigorous evaluation and validation procedures were employed to assess the model's performance, providing confidence in its ability to generalize its learning to new and unseen samples. Therefore, magnetites can shed light on its origins and evolution.

2. Literature Review

2.1. The Classical Approaches

Authors, e.g., [7], have provided the classical approach that aims to distinguish between several types of ore deposits, including iron-oxide–copper–gold (IOCG), Kiruna-apatite–magnetite, banded iron formation (BIF), porphyry Cu, Fe-Cu skarn, Fe-Ti, V, Cr, Ni-Cu-PGE, Cu-Zn-Pb volcanogenic massive sulphide (VMS), and Archean Au-Cu porphyry and Opemiska Cu veins. Their work is based on the use of the trace element fingerprinting of magnetite to classify each deposit type. The authors proposed a three-step approach to classify various mineral deposit types based on the trace elements from magnetites. The method was designed to sequentially distinguish different deposit types.

- Nickel deposits: The first step in the approach was to identify magnetite samples associated with Ni deposits. These deposits have magnetite with high Ni and Cu concentrations. By plotting the concentrations of Ni + Cr vs. Si + Mg, the authors were able to discriminate Ni-Cu-PGE deposits from other deposit types.
- Volcanogenic massive sulphide (VMS) deposits: The second step aimed to distinguish VMS deposits from the remaining deposit types. VMS deposits are characterized by magnetite with high Zn concentrations. By plotting the concentrations of Al/(Zn + Ca) vs. Cu/(Si + Ca), the authors were able to discriminate VMS deposits from the other types.
- Other deposit types: The final step involved differentiating the remaining deposit types, including IOA, IOCG, Algoma-type iron formations, porphyry Cu deposits, and skarn deposits.
- This was achieved by analysing the trace element concentrations in magnetite, such as Ti, V, Cr, Mn, Co, Cu, Ge, Y, Zr, Nb, Mo, Sn, Hf, Ta, W, Pb, Th, and U. By plotting Ca + Al + Mn vs. Ti + V elements, the authors were able to create discriminant diagrams that revealed distinct clusters or trends for different deposit types.
- Other authors [9,10] highlight the systematic variations in the concentration of minor and trace elements in the composition of hydrothermal magnetite from different deposit types and igneous magnetite from mineralized and barren host rocks. The primary factors governing magnetite compositional variations include temperature, fluid composition, oxygen and sulphur fugacity, silicate and sulphide activity, host rock buffering, re-equilibration processes, and intrinsic crystallographic controls. Key discriminator elements such as Mg, Al, Ti, V, Cr, Mn, Co, Ni, Zn, Ga, and Sn play a vital role in distinguishing magnetite from different deposit types and host rocks.

The authors propose that a combination of discriminant plots, such as Ti + V vs. Al + Mn and box and whisker plots for magnetite elements, can effectively differentiate between magnetite from BIFs, Ag-Pb-Zn veins, porphyry and associated skarn deposits, Climax-Mo deposits, and barren igneous host rocks. They note that low levels of Al, Ti, V, Cr, Co, Zn, and Ga are particularly diagnostic for magnetite from BIFs, Mg-skarn, and Ag-Pb-Zn veins. Additionally, while igneous and hydrothermal magnetite from porphyry deposits share similar compositional patterns, they can be differentiated based on Cr, Co, Ni, and Ga concentrations, with Ga concentrations being consistently higher in igneous magnetite than in hydrothermal occurrences. This research supports and refines the classification scheme proposed by [7] for distinguishing magnetite from various types of mineral deposits.

2.2. The Machine Learning Approach

In a study by [16], a random forest (RF) classifier was developed to identify ore deposit types based on the concentrations of Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Tl, and Pb in pyrite. The classifier demonstrated a high accuracy for both test and blind test data, achieving overall accuracies of 94.5% and 93.9%, respectively, when excluding inconclusive analyses. The authors concluded that random forest classifiers derived from the chemistry of individual minerals hold potential as a useful geochemical exploration tool. However, they stress that this should be seen as a preliminary positive result, and additional pyrite varieties must be incorporated into the classifier before widespread application in mineral exploration. Moreover, the authors emphasize that the classifier should be regarded as one of many tools rather than a single stand-alone classification method.

Other authors [13] refer to an RF classifier being employed to classify IOCG (iron-oxide-copper-gold) and IOA (iron-oxide-apatite) deposits based on the chemical composition of iron oxides measured using an Electron Probe Microanalyzer (EPMA) and LA-ICP-MS. The study found that despite the complex and uneven geochemical composition of iron oxides, IOCG and IOA deposits can be distinguished. The authors emphasized that the training set and testing datasets should not be divided randomly; instead, they should be divided according to the deposit.

The RF model performance using the LA-ICP-MS dataset was found to be superior to the classification model based on EPMA data, indicating that incorporating more geochemical variables and higher-quality data is beneficial in distinguishing IOA and IOCG deposits. Moreover, the study identified key elements for discriminating IOCG from IOA deposits: V, Mg, and Mn in the EPMA data, and Si, Mg, and V in the LA-ICP-MS data.

In [14], the authors employed machine learning techniques to analyse high-Ti magnetite samples from various deposit types, including igneous, porphyry, IOA, and IOCG deposits. The primary objective was to investigate the controlling elements of magnetite compositions and elucidate the origin of IOA magnetite. The findings demonstrate that machine learning techniques, when applied correctly, are highly effective in distinguishing high-Ti magnetite from different origins. By appropriately removing noise elements, the interpretability of machine learning results can be significantly enhanced.

The results reveal that magnetite from IOA deposits exhibits elemental characteristics similar to high-temperature hydrothermal magnetite from IOCG and porphyry systems, while being distinctly different from igneous magnetite. This suggests that IOA magnetite is most likely of a hydrothermal origin. The study also proposes a new discriminant diagram for high-Ti magnetite, $\lg(\text{Al}) + \lg(\text{Ti}) + \lg(\text{V})$ versus $\lg(\text{Mn})/[\lg(\text{Co}) + \lg(\text{Mg})]$. The support vector machine (SVM) method indicated that this new diagram has an accuracy of 97.6%. This new discriminant diagram effectively captures the characteristics of data distribution observed in dimensionality reduction techniques, such as PCA and t-SNE projections, thereby offering a valuable tool for distinguishing high-Ti magnetite from different origins.

Authors [15] used an extensive dataset with more than 17 k observations from 303 different deposit types including BIF, Fe-Ti, IOCG, IOA, Ni-Cu-PGE, porphyry, VMS, skarn, and V. These authors tested the performance of three machine learning models, naïve Bayes, k-nearest neighbour, and random forest, and concluded that random forest is the best model for most of the deposit types and that it can be used successfully in exploration campaigns to identify the different mineral deposit types.

2.3. The Portuguese OMZ Case Study

The case study comprises LA-ICP-MS data of magnetite trace element composition from six iron deposits located in the Portuguese Ossa-Morena Zone (Figure 1a). The selected deposits belong to the Montemor-o-Novo—Ficalho metallogenic province [17], a subdivision of the Évora-Aracena belt [18].

- Montemor-o-Novo Iron Complex

Three of the selected deposits, Monges, Vale da Arca, and Serrinha deposits, belong to the volcanogenic-exhalative Montemor-o-Novo Iron Complex [5]. The MIC comprises several iron deposits (Figure 1b), many of which were mined during the early 20th century, and their main metallogenic input is associated with an early–middle Cambrian intracontinental rifting stage (Variscan cycle; [19]). The iron ores are mainly hosted by calcite–dolomite rocks from the Monfurado Formation [20], which were metasomatized in the late stages of the Variscan cycle [21]. Recent works proposed a combined volcanogenic-exhalative (SEDEX-VMS) origin for the MIC deposits [17,22], with ore deposition being associated with the input of hydrothermal fluids during the early stages of the oceanization of the Rheic.

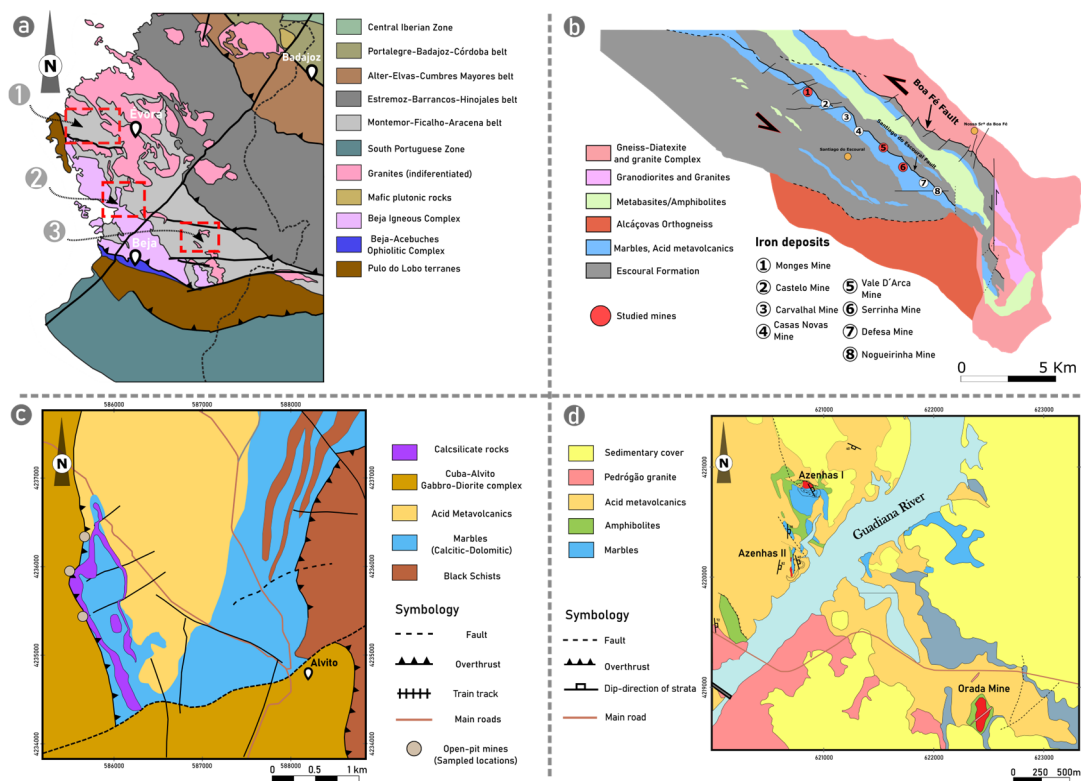


Figure 1. Geological mapping of the selected deposits for this study. (a). Representative geotectonic arrangement of Ossa-Morena Zone [23] with location of the Fe deposits selected for this study (red rectangles; adapted from [6,24]). (b). Geological map of the Montemor-o-Novo Shear Zone (adapted from [5,20] with the location of the main iron deposits that integrate with the Montemor-o-Novo Iron Complex [25,26]. The circles filled in red represent the deposits selected for this study. (c). Geological map of the Alvito area (adapted from [6,27–29]) with representation of the open-pit mining location in which magnetite samples were collected and further examined with LA-ICP-MS [5]. (d). Geological map of the Azenhas (Azenhas I and Azenhas II) and Orada deposits [6,30].

- **Alvito deposit**

The Alvito deposit (Figure 1c) is a Fe-skarn deposit associated with the emplacement of a gabbro-diorite suite, part of the Beja Igneous Complex (Figure 1a). The emplacement of this magmatic body started at $350 \text{ Ma} \pm 5 \text{ Ma}$ [31], and in the Alvito area, it contacts carbonate rocks and generates an exoskarn that hosts most of the massive magnetite. Ore deposition is attributed to the retrograde stages of metasomatism [5], and magnetite is found in diopside–hedenbergite-rich sections.

- **Azenhas–Orada deposits**

The Azenhas–Orada deposits (Figure 1d) are a complex of iron orebodies in which mining took place up until the second half of the 20th century. The massive orebodies are mostly hosted in metasomatized allochthonous amphibolites [17,18,32], which are tectonically imbricated over Cambrian carbonates. The genesis of these deposits is discussed, and previous authors have attributed their genesis to a prolonged thermal gradient associated with the tectonic imbrication. This tectonic arrangement is thought to have favoured the deposition of magnetite related to the metasomatism of the amphibolites. Recently, a magnetite trace element analysis revealed that the genesis of these deposits might be associated with higher-temperature magmatic processes [5].

3. Materials and Methods

Considering the previous studies, we envisaged to apply a supervised classification approach to a set of iron deposits from the OMZ, in order to verify the previous conclusions and methodologies [5,6]. To conduct our analysis, we utilized the R programming language and a selection of well-established packages designed for implementing machine learning techniques. The ‘caret’ package [33] serves as the foundation for our analysis, providing a streamlined interface for data pre-processing, model training, and performance evaluation. For the classification tasks, we employed the ‘randomForest’, ‘naivebayes’, ‘nnet’, and ‘e1071’ packages [34–37], which allow us to implement the proposed machine learning classifiers. By harnessing the capabilities of these packages, we efficiently explore the various models, compare their performance, and ultimately identify the most suitable approach for our research objectives.

3.1. Datasets and Pre-Processing of Data

Literature data were collected to create a Reference Dataset with data from LA-ICP-MS analyses of magnetite from Banded Iron Formation (BIF) deposits [38]; iron oxide, copper, and gold (IOCG) deposits [39]; porphyry-type deposits [40]; Kiruna-type iron-oxide-apatite (IOA) deposits [41]; skarn deposits [42]; and volcanogenic massive sulphide (VMS) from the Izok Lake VMS [43,44].

This Reference Dataset for the self-evaluation process was divided into two groups: (i) a random sample of the Reference Dataset, named the Train Dataset and containing 70% of the data. This group includes 190 magnetite measures that are used for training the different ML models; (ii) the remaining data, the Test Dataset, corresponds to the remaining 30% of the Reference Dataset, i.e., 81 magnetite observations that are used for evaluating the results from the ML classifier.

The Case Study Dataset is the data collected from [5,6]. For training the ML models for the Case Study Dataset, all the data from the Reference Dataset were used, i.e., the 271 magnetite values. For all these datasets, and accordingly to what was previously defined in the bibliography, nine elements were used (Mg, Al, Ti, Mn, V, Co, Ni, Zn, Ga).

To circumvent the limitation of censored data in such large datasets, the substitution of the analysis below the detection limit of the equipment was performed using the log-ratio Expectation–Maximisation (EM) algorithm through the *lrEM* function in a *zCompositions* package [45] in the R programming language environment. This package allows for the inspection of the censored values using the *zPatterns* function so that the variables with over 40% missing values can be excluded from the dataset [46].

3.2. Dimension Reduction

A Principal Component Analysis (PCA) is a dimensionality reduction technique that, when applied to the analysis of magnetite composition, can provide insights for the discrimination of ore deposits [11]. PCA can help visualize and interpret complex, multidimensional geochemical data more effectively. The elemental concentrations of magnetite are transformed into a new set of uncorrelated variables, the principal components (PCs). These PCs are linear combinations of the original variables and are ordered in such a way that the first principal component (PC1) captures the largest variance in the dataset, followed by the second principal component (PC2), and so on. By retaining only a few PCs that explain a significant portion of the variance, PCA effectively reduces the dimensionality of the data while preserving the essential information and patterns.

When using magnetite composition for ore deposit discrimination, PCA can help reveal geochemical trends and associations between different deposit types based on the major PCs [47]. By plotting the scores of the samples on the first few PCs, it is possible to visualize the clustering and separation of different deposit types within the reduced-dimensional space. This visualization helps in understanding the relationships between magnetite compositions and ore deposit types, therefore contributing to strengthen the geological models applied to certain geological targets.

3.3. Machine Learning

Random forest (RF) is an ensemble learning method that operates by constructing multiple decision trees during a training phase. The output of the random forest model is determined by aggregating the predictions from individual trees, which helps to minimize overfitting and improves the model's generalization. This method is particularly effective in handling large datasets with a high number of features, as it can efficiently handle missing values, outliers, and imbalanced data [13,16,48,49]. For this classification, 500 decision trees were defined. The cross-validation method for resampling involved a 10-fold partition of the data.

Naïve Bayes (NB) is a probabilistic machine learning algorithm based on the application of Bayes' theorem with the assumption of independence between features. This method is particularly effective for handling large datasets efficiently [50]. Despite the simplifying assumption of feature independence, naïve Bayes often performs well in practice, especially when the independence assumption is approximately true or when the model's purpose is to rank rather than predict absolute probabilities.

Support vector machines (SVM) are a class of supervised learning algorithms used for both classification and regression tasks. The core idea behind SVM is to find the optimal hyperplane that best separates the data points into different classes, with the largest possible margin between the classes. In the case of non-linearly separable data, SVM employs kernel functions, such as the radial basis function (RBF) used in this study. This function transforms the data into a higher-dimensional space where a linear separation becomes possible. SVM is known for its robustness against overfitting and ability to handle high-dimensional data effectively. However, the choice of an appropriate kernel function and tuning of the hyperparameters can be challenging, and the algorithm can be computationally expensive for large datasets. SVM has been effectively used in the classification of ore deposits using LA-ICP-MS data from magnetite [14], as well as in identifying areas for potential exploration [51,52].

Multinomial logistic regression (MLR), also known as SoftMax Regression or a MaxEnt Classifier, is an extension of logistic regression that allows for the prediction of multiple classes rather than just binary outcomes. This method is particularly useful for multi-class classification problems, where the target variable has more than two possible categories. Multinomial logistic regression operates by estimating the probabilities of an input instance belonging to each class using the SoftMax function, which normalizes the input's linear combination into a probability distribution over the classes. The predicted class is then determined by selecting the class with the highest probability. While still maintaining the simplicity and interpretability of logistic regression, multinomial logistic regression enables the model to handle more complex classification problems effectively.

K-nearest neighbour (KNN) is a non-parametric, instance-based learning algorithm used for both classification and regression tasks. It operates by comparing a new, unlabelled instance with a set of labelled instances, determining the k-nearest instances according to a distance metric, and then predicting the label based on the majority class or a weighted average of the neighbours' labels. KNN is known for its simplicity, interpretability, and ability to adapt to new data, but its performance can be negatively impacted by high-dimensional data or large datasets due to the curse of dimensionality and increased computational cost [53].

Artificial neural networks (ANNs) are a family of machine learning models inspired by the structure and function of biological neural networks. An ANN consists of interconnected layers of artificial neurons, which process input data and adjust their weights based on the error between the predicted output and the actual target values. ANNs are highly flexible and can learn complex, non-linear relationships, making them suitable for a wide range of tasks; however, they can be computationally expensive and often require a large amount of data to achieve an optimal performance [52,54,55]. In the case of our study, the 'nnet' package was used and tested with 1 to 15 hidden layers and a feedforward neural

network with a sigmoid activation function. The best performing model was selected for classification purposes.

4. Results

4.1. Exploratory Data Analysis

The first approach to understand the data *behaviour* is to use descriptive statistics, namely the central tendency descriptors. For the Reference Dataset, the mean, standard deviation (sd), skewness (skew), and kurtosis were calculated for the six types of ore deposits. The mean and the standard deviation provide an indication of the values involved in each case and its variation, whereas the skewness provides a sense of a normal or asymmetric tendency of the data and the kurtosis represents the data distribution as being flat or wedged shaped.

Table 1 displays these results grouped by the type of deposit.

Table 1. Descriptors for the Reference dataset.

	BIF					IOA					IOCG							
	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis			
Al	39	640.64	529.01	1.3	1.3	55	4308.84	2894.45	3.7	17.9	34	316.26	110.14	0.7	−0.3			
Mg	39	944.89	948.55	1.0	0.3	55	5570.18	2738.01	0.5	−0.4	34	46.14	57.92	1.6	1.8			
Ti	39	53.20	43.51	1.1	0.9	55	11,800.67	18,182.72	4.1	17.8	34	82.88	89.16	3.9	15.2			
V	39	6.61	8.27	2.5	6.1	55	2569.07	765.08	1.6	4.0	34	654.38	548.12	0.8	−1.2			
Mn	39	198.67	404.15	2.5	5.6	55	619.30	389.83	2.0	5.8	34	43.36	45.84	1.9	2.5			
Co	39	2.88	2.25	0.4	−1.3	55	112.52	21.32	−1.5	2.0	34	76.73	17.70	−1.7	2.4			
Ni	39	8.09	3.38	2.0	5.2	55	319.11	61.25	−0.1	1.2	34	168.30	101.21	0.6	−1.0			
Zn	39	39.26	19.11	0.8	0.7	55	99.46	34.66	0.2	−0.6	34	5.45	3.23	1.2	0.7			
Ga	39	2.65	1.08	0.2	−1.1	55	26.90	8.18	−0.4	−1.3	34	15.62	8.61	0.7	−1.0			
		Porphyry						Skarn						VMS				
	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis			
Al	68	3010.65	2072.08	0.6	−0.2	52	6121.81	2168.93	−0.3	−0.5	23	1108.97	1274.49	1.2	−0.3			
Mg	68	1028.38	962.33	1.1	0.6	52	5886.31	2603.75	1.2	0.3	23	157.09	337.43	2.8	7.8			
Ti	68	5094.00	6555.82	1.6	1.7	52	984.92	467.10	−0.5	−1.0	23	601.96	875.89	2.3	4.8			
V	68	927.73	777.27	1.4	2.0	52	193.66	52.21	−1.1	1.5	23	916.87	1096.34	0.9	−0.7			
Mn	68	2048.01	2353.73	2.5	6.5	52	2916.97	1616.78	1.5	1.7	23	801.13	587.59	0.7	−1.2			
Co	68	44.60	27.33	0.4	0.0	52	62.09	27.61	2.3	5.2	23	18.58	17.57	1.0	−0.1			
Ni	68	55.19	73.60	3.5	13.1	52	6.98	2.66	−1.0	0.0	23	20.78	28.52	1.5	0.8			
Zn	68	647.48	513.99	1.2	0.7	52	283.40	129.73	1.1	0.4	23	151.70	219.68	2.9	8.6			
Ga	68	57.64	29.11	−0.5	−0.6	52	21.17	8.58	−0.5	−1.4	23	49.99	46.83	1.5	1.4			

Regarding the mean and standard deviation, it is notorious that values are mostly different for different elements and for each deposit type; nevertheless, there are some overlapping values, e.g., Ti in porphyry ($\bar{x} = 2057.78$, $\sigma = 1751.09$) and skarn ($\bar{x} = 2111.46$, $\sigma = 483.71$), or Zn in BIF ($\bar{x} = 39.26$, $\sigma = 19.11$) and VMS ($\bar{x} = 49.80$, $\sigma = 17.23$).

The skewness values in the majority of the deposit types are weakly positive, indicating a good Gaussian distribution. The kurtosis values have a moderate positive tendency, indicating a slightly wedged distribution pattern, hence with a restricted variation among the central values. The exception to this assertion is the VMS deposits with a high skewness and kurtosis for the Al, Mg, and Ni elements. A second group of outstanding values are in IOA deposits, the Al and Ti in IOCG, all exhibiting kurtosis values higher than 10.

The results of the exploratory data analysis for the Case Study Dataset are presented in Table 2.

Table 2. Descriptors for the Case Study dataset.

	Alvito					Azenhas					Monges				
	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis
Al	159	9851.18	2502.80	−0.1	0.5	22	2002.19	1302.13	0.7	−0.7	98	914.55	1134.41	4.7	29.8
Mg	159	3730.98	1998.24	1.1	5.0	22	8092.75	3016.74	1.0	0.4	98	2945.82	2913.96	5.5	41.4
Ti	159	2057.78	1751.09	1.0	−0.6	22	2111.46	483.71	1.2	0.3	98	569.98	212.23	0.8	−0.3
V	159	70.43	28.93	−0.3	−0.9	22	129.94	29.47	−0.3	−1.1	98	37.74	17.63	0.9	−0.2
Mn	159	2149.33	568.79	0.7	0.8	22	4827.21	681.53	0.2	−1.1	98	6927.25	4389.85	0.5	−1.1
Co	159	85.93	24.92	0.2	0.1	22	52.19	13.28	0.1	−0.7	98	0.37	0.36	0.8	−0.8
Ni	159	40.44	29.47	0.1	−1.3	22	17.01	4.56	0.3	−1.4	98	1.35	1.33	5.2	34.2
Zn	159	523.72	204.85	0.2	−0.5	22	658.76	244.29	−0.2	−1.0	98	49.80	17.23	−0.1	−1.0
Ga	159	39.62	12.29	0.7	1.5	22	12.78	8.07	0.2	−1.9	98	6.21	2.45	0.4	−0.8
	Orada					Serrinha					Vale da Arca				
	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis	n	mean	sd	skew	kurtosis
Al	35	2772.54	685.00	−0.2	−0.8	45	715.37	153.65	1.0	−0.1	12	67.14	49.04	0.4	−1.4
Mg	35	2597.43	589.30	1.9	4.6	45	4427.55	1453.70	2.0	4.5	12	2254.57	512.76	−0.2	−0.9
Ti	35	144.14	10.34	0.0	0.1	45	69.27	46.96	0.6	−1.3	12	142.46	42.48	−1.1	0.9
V	35	44.18	3.74	−0.8	3.2	45	8.95	10.25	1.2	−0.1	12	132.05	38.55	−1.8	2.7
Mn	35	497.61	54.71	−0.1	0.9	45	11,267.92	1178.06	0.2	0.0	12	7250.09	687.79	0.0	−0.9
Co	35	28.97	2.33	−1.3	4.3	45	0.53	0.06	1.1	2.7	12	0.03	0.01	2.6	5.5
Ni	35	24.40	2.42	−0.7	1.3	45	1.14	0.49	2.1	6.9	12	15.84	14.88	1.0	−0.7
Zn	35	68.82	12.62	0.3	−0.6	45	46.59	8.25	0.6	−0.4	12	109.27	15.32	−0.2	−0.9
Ga	35	8.38	0.85	−0.7	1.9	45	0.62	0.17	0.7	−0.7	12	3.47	0.52	−0.1	−1.2

Considering the Case Study Dataset, as expected from the recognized deposit types, Alvito and Azenhas deposits group together for most of the elements and differ considerably from the other deposits; this is more noticeable by displaying higher values for Al, Mg, Ti, Co, Ni, Zn, and Ga. Contrarily, Mn is lower in these deposits compared with the other ones. The Orada deposit stands out with distinctive trends; in some cases, with compositional values similar to the Alvito and Azenhas deposits and, at other times, with mean values more alike the other types of deposits.

The skewness in most of the case studies is weakly positive, with the exception of the Orada deposit, where it is weakly negative. Standing out from this trend is the Monges deposit, where Al, Mg, and Ni have strong positive skewness and kurtosis values.

4.2. The Classical Approach

The mineral deposit classification based on the discriminant diagrams for the Reference Dataset is plotted in Figure 2. The VMS samples are not meant to be plotted in this diagram and, hence, cross several of the defined fields (see [7]). Porphyry data coincide partly with the IOA. The IOCG data from the Reference Dataset fall outside the defined field, presenting lower than expected values of Al + Mn and somehow inside the BIF field. On the other hand, the BIF data are outside the identified fields, falling in lower than expected Ti + V areas.

This demonstrates the difficulty of using these diagrams, as was previously stated by authors in, for example [9,14].

Despite this difficulty, it is also clear that these diagrams can easily separate magnetites from some of the main types of deposits, but others such as porphyry, VMS, or IOA have overlying areas in these diagrams.

The data from the case study, plotted in the diagram [7], are presented in Figure 3. According to this diagram, all data are plotted, and thus classified, in the skarn-type field, essentially with varying amounts of the Ti + V component.

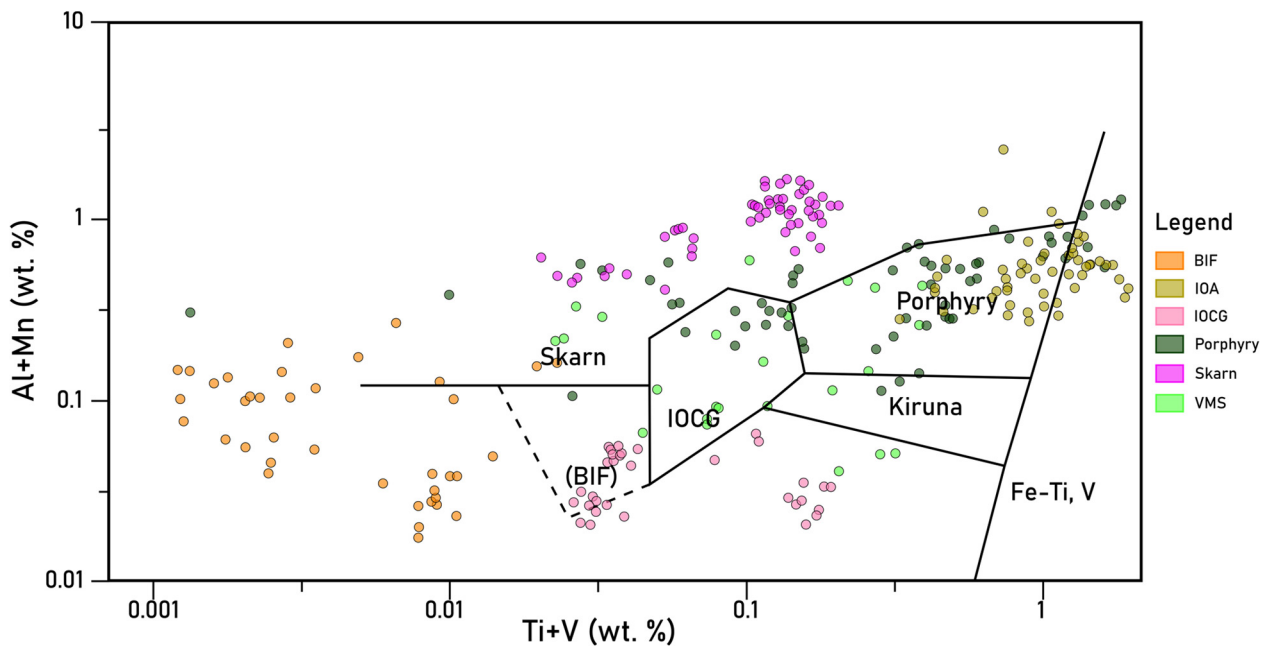


Figure 2. The Reference Dataset plotted in the space of the [7] diagram. The reference dataset was gathered from the following authors: BIF [38]; IOA [41]; IOCG [39]; Porphyry [40]; Skarn [42]; VMS [43].

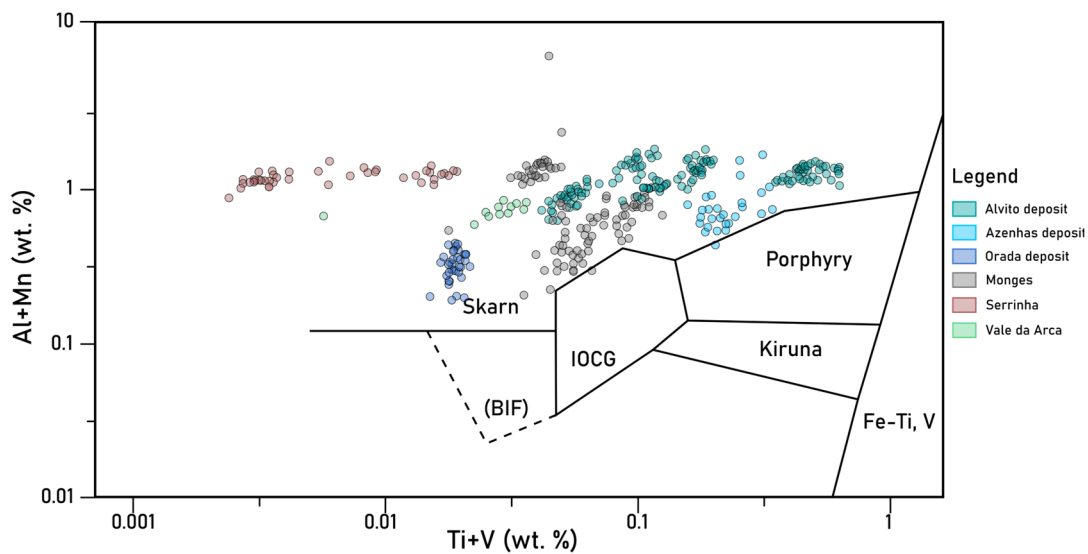


Figure 3. The Case Study Dataset plotted in the [7] diagram.

It is readily recognized that the two types of deposits from the case study of the Portuguese OMZ are not easily distinguishable using this diagram.

4.3. Dimension Reduction (PCA)

The PCA analysis applied to the Reference Dataset shows that PC1 explains 32.48% of the variability and is dominated by the behaviour of Co, V, and Ni. The PC2 component explains 26.26% of the variability and is dominated by the elements Zn, Mn, and Ga.

The equation derived from the three main elements that influence PC1 is

$$0.49 \times Co + 0.46 \times Ni + 0.46 \times V \tag{1}$$

and for PC2, the equation is

$$-0.57 \times \text{Zn} - 0.53 \times \text{Mn} - 0.37 \times \text{Ga} \tag{2}$$

These results enhance the role of the combined elements Co, Ni, and V for explaining the most significant part of the variability (PC1) and the combined Zn, Mn, and Ga for explaining another part (PC2). A PC1–PC2 plot with ellipses corresponding to 75% of the data for each group is presented in Figure 4. A good discrimination between BIF, IOCG, skarn, and IOA deposits is possible using the PC1–PC2 diagram, whereas the porphyry and VMS show a significant overlap between them and the other deposits.

Reference Dataset

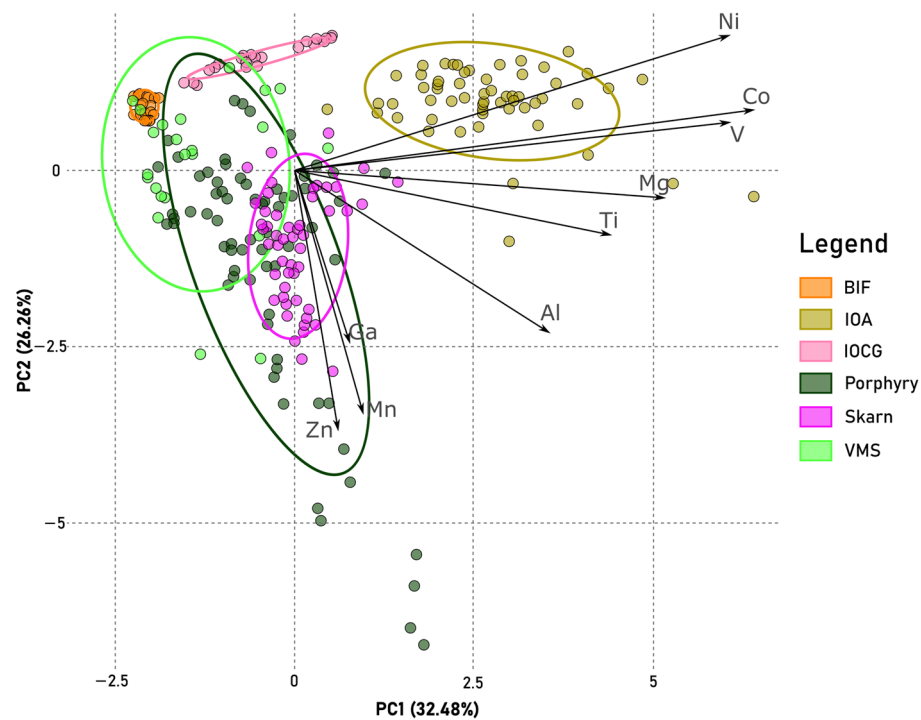


Figure 4. The Reference Dataset plotted in the space of the [7] diagram. The reference dataset was gathered from the following authors: BIF [38]; IOA [41]; IOCG [39]; Porphyry [40]; Skarn [42]; VMS [43].

The PCA analysis applied to the Case Study Dataset reveals a good discrimination between all the studied deposits (Figure 5). PC1 explains most of the variance with 53.64%, whereas PC2 explains less variability (17.44%). An exception to this is the Monges deposit in which the data overlap to some extent with the Serrinha and Vale da Arca deposits.

The equation of the three most significant variables for PC1 is

$$0.41 \times \text{Zn} + 0.40 \times \text{Co} + 0.39 \times \text{Al} \tag{3}$$

and for PC2:

$$0.65 \times \text{Mg} + 0.37 \times \text{Mn} + 0.30 \times \text{Ti} \tag{4}$$

In this case, there is a clear opposition of the Zn, Co, and Al group with the Mg, Mn, and Ti group.

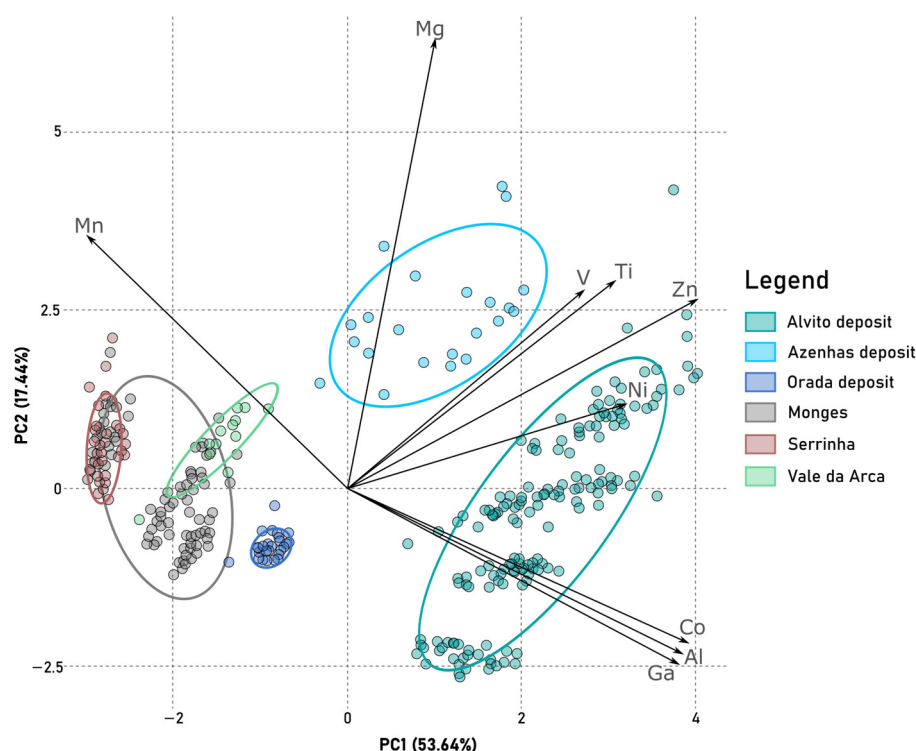


Figure 5. The PC1–PC2 plot for the Case Study Dataset.

4.4. Machine Learning

The results of the different classifiers are better understood when analysing the confusion matrices that are an indicator of the performance of the model. The criteria to consider a correct classification of the deposits were adopted from [16], where accuracy is defined as a measure of the quality of the results.

One common measure of this accuracy is the F1-score that is defined as

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

where precision = $TP / (TP + FP)$ and recall = $TP / (TP + FN)$, in that TP are the true positives, FP are the false positives, and FN are the false negatives.

Therefore, accuracies above 65% are considered as a correct classification, those between 55% and 65% are regarded as inconclusive, and those below 55% are deemed incorrect.

The permutation importance value is used to identify the variable importance in each model [55]. This parameter measures the change in model performance after shuffling the values of a given variable. This calculation is performed using the `varImp()` function from the ‘caret’ package [33]. If the model’s performance decreases, it indicates that the feature is important for making accurate predictions. If the performance remains the same or improves, it suggests that the feature is less important or even harmful to the model’s predictive ability. A positive value indicates that the model’s performance decreased after the feature was shuffled; hence, the feature is important for the model to make accurate predictions. The larger the positive value, the more important the feature. A negative value indicates that the model’s performance improved or remained the same after the feature was shuffled.

The F1-score is another metric used to evaluate the performance of a machine learning model. It combines the precision and recall metrics into a single score. Precision measures the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positive predictions among all actual positive samples. The F1-score is the harmonic mean of precision and recall. The F1-score ranges from 0 to 1,

with 1 being the best possible score. A high F1-score indicates that the model has both a high precision and high recall, which means that it can correctly identify positive samples and minimize false positives and false negatives. A low F1-score indicates that the model is not performing well, either because it has a low precision, low recall, or both.

Both the permutation importance value and F1-score are only applicable to the Reference Dataset and are considered to evaluate the performance of the models.

4.4.1. Reference Dataset Modelling

As stated above, the Reference Dataset was divided into the Train Dataset and the Test Dataset. The results from classifying the Test Dataset after training the model with the Train Dataset provide the following results.

- Random forest (RF)

The RF model was run with 500 trees, which is a good compromise between computation time and accuracy. The results from applying the RF model are presented as a confusion matrix in Table S1 (Supplementary Material). There is a 100% true positive value prediction for IOCG and skarn deposit types, with a good precision for porphyry, IOA, and BIF deposits. VMS deposit identification has the worst precision value.

To illustrate one of the decision trees created for the RF model, Figure 6 displays an example of one of the trees; in this case, the one with a lower number of branches. Notice that not all the chemical elements are used in this tree, in which only Ti, Ni, Mg, Ga, V, and Mn values correspond to decision nodes used for classification. In this case, the most relevant elements are Ti and Ni.

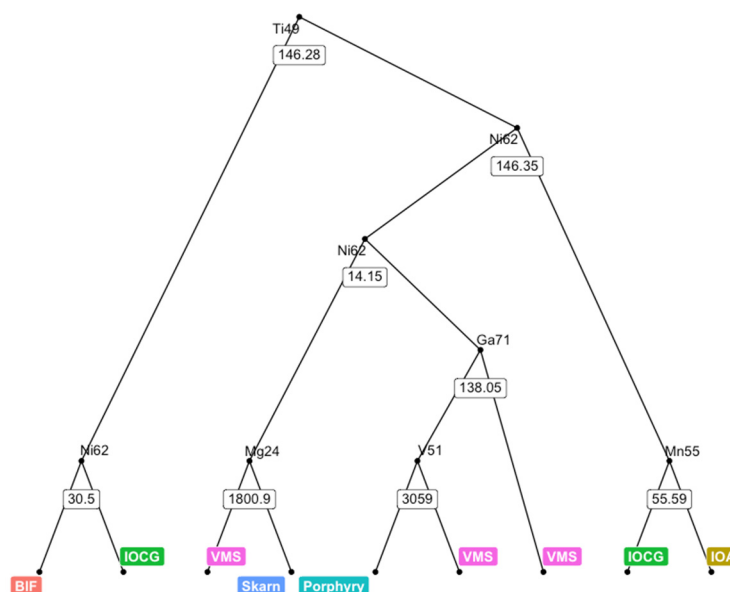


Figure 6. An example of a random decision tree created from the Reference Dataset.

Regarding the variable importance, measured with the permutation method, Mg has a higher value (0.25), followed by Zn (0.04) and Ni (0.02); all other variables have a lower importance (<0.01) in this model.

The overall accuracy considering the weighted F1-score for this model is 92%.

- Naïve Bayes (NB)

The results using the NB model are displayed in a confusion matrix presented in Table S2 (Supplementary Material). With this classifier model, IOA, IOCG, and skarn deposits have a perfect true positive prediction. BIF and porphyry have a good precision, whereas VMS deposit types have lower values.

The variable importance calculated using the permutation method points to a higher importance connected to Ni (0.04), followed by Ti (0.02) and V (0.02). All the other variables show a limited importance considering this model prediction.

The F1-score with the overall weighted accuracy for the NB model is 89%.

- Support Vector Machine (SVM)

The SVM model classification results are represented by a confusion matrix in Table S3 (Supplementary Material). This classifier has a perfect classification for IOA, IOCG, and skarn deposits; BIF and porphyry have correct classifications and VMS deposits have incorrect classification results.

The permutation importance of the different variables indicates that Ga (0.15), Al (0.12), Co (0.11), and V (0.10) are the most meaningful variables. Nevertheless, all the others, with the exception of Ti, have values higher than 0.04.

The weighted accuracy using the F1-score value is 83%.

- Multinomial Logistic Regression (MLR)

The results using the MLR model, expressed as a confusion matrix, are presented in Table S4 (Supplementary Material). This classifier has perfect classification results for IOA and IOCG, presenting good results for all other types of deposits.

The permutation importance values indicate that Co (0.33), Mg (0.32), Ga (0.29), and Ni (0.28) have higher values and therefore the most relevant role in defining the model. Nevertheless, the other variables have significant values, where Ti has the lowest significance (0.07).

The overall weighted accuracy measured using the F1-score of this model is 88%.

- K-Nearest Neighbour (KNN)

The results using the KNN classifier model are presented as a confusion matrix in Table S5 (Supplementary Material). In this case, the IOA deposits have a perfect classification, whereas BIF and skarn have a precision higher than 80%. IOCG, porphyry, and VMS have correct classification but have lower values.

Regarding the permutation importance of the different variables, the results indicate that Mg (0.34), Al (0.31), and Ti (0.28) have higher values, whereas Ni, Co, Zn, and Ga have an importance lower than 0.02.

The F1-score for this model is 77%.

- Artificial Neural Network (ANN)

The results using the ANN model approach are presented as a confusion matrix in Table S6 (Supplementary Material). This is the worst performing model, where none of the deposit types receive a perfect classification. Nevertheless, IOA, skarn, BIF, and IOCG have a precision equal to or higher than 80%, porphyry deposit types have correct classification but with a low precision value, and VMS classification is inadequate.

The permutation importance using this model indicates that Mg (0.33), Mn (0.31), V (0.30), Al (0.26), and Ti (0.26) have a higher importance. The other chemical elements have a very low importance in the model classification.

The F1-score using the ANN model is the lowest with a value of 76%.

The best performing models in decreasing order of F1-score are random forest (93%), naïve Bayes (89%), multinomial logistic regression (88%), support vector machine (83%), k-nearest neighbour (77%), and artificial neural networks (76%).

4.4.2. Case Study Dataset Modelling

The full Reference Dataset, with 271 observations, was used to train the machine learning (ML) models to be applied in the Case Study Dataset. These models were then used to predict the classification of the Ossa-Morena deposits. The cross tables between the OMZ deposit and the proposed classification with the ML model prove to be the best method to evaluate the obtained results.

A measure of the assertiveness (Assert.) of each model is used to quantify the percentage of values that fall in the same model type. Higher values indicate that most of the observations fall in the same deposit type. Notice that this does not mean an accurate classification of the deposit and should be interpreted as an indicator of the homogeneity of the classification. These values are presented in the last column of the following tables. In these tables, the highlighted values refer to the most suggested classification using the model. In some cases, there is no clear decision, indicating two possible classifications; for example, in Table S9, the Monges deposit falls equally in the BIF and VMS types with 40 observations proposed for each deposit type.

- Random Forest classification model

The RF model for Alvito and Azenhas deposits indicates a skarn deposit type (Table S7 in Supplementary Material), although in Alvito, a high number of observations fall in the porphyry deposit type. Orada is classified as the VMS deposit type. Monges, Serrinha, and Vale da Arca are considered as BIF deposit types; however, Monges observations are equally classified as the VMS deposit type.

A better assertiveness is achieved for the Azenhas deposit, as a skarn, and Serrinha and Vale da Arca as a BIF.

- Naïve Bayes classification model

The NB model classification for the case study is somewhat similar to the RF mode (Table S8 in Supplementary Material). However, Alvito and Azenhas fall almost equally between porphyry and skarn deposit types. The Orada and Monges deposits are mostly classified as VMS deposits. Serrinha and Vale da Arca are classified as BIF deposit types.

Concerning assertiveness, Orada, Serrinha, and Vale da Arca are the ones that perform better.

- Support Vector Machine classification model

The SVM model classification is presented in Table S9 in Supplementary Material. Alvito and Azenhas are classified as skarn deposits, although Azenhas has many observations considered as porphyry. Orada is classified as a BIF deposit, whereas Monges, Serrinha, and Vale da Arca are classified as porphyry deposit types.

Concerning assertiveness, Alvito, Serrinha, and Vale da Arca are the deposits that perform better.

- Multinomial Logistic Regression classification model

The MLR model classification is shown in Table S10 in Supplementary Material. This model indicates that Alvito and Azenhas fall in the skarn deposit type. The Orada deposit is considered as a BIF deposit. Monges, Serrinha, and Vale da Arca are classified as the porphyry deposit type.

Concerning assertiveness, Alvito, Azenhas, Orada, and Vale da Arca are the deposits that perform better. This model is also the one that, overall, shows better assertiveness.

- K-Nearest Neighbour classification model

The KNN model classification is shown in Table S11 in Supplementary Material. Alvito and Azenhas are classified as skarn deposits. Orada falls both in the BIF and porphyry deposit type. Monges and Vale da Arca are mostly classified as porphyry, whereas Serrinha is defined as a skarn deposit.

Concerning assertiveness, Alvito and Vale da Arca are the deposits that perform better.

- Artificial Neural Network classification model

The ANN model classification is shown in Table S12 in Supplementary Material. Alvito is classified as a skarn deposit, but also falls in the BIF classification jointly with Azenhas. Orada is classified as a skarn deposit, whereas Monges is classified as being porphyry and Serrinha and Vale da Arca are classified as VMS deposits.

Concerning assertiveness, Orada, Serrinha, and Vale da Arca are the deposits where this model performs better.

5. Discussion

5.1. Dimension Reduction Using PCA

The dimension reduction approach, namely the PCA, is demonstrated to be a good tool for understanding the behaviour of the different chemical elements in the system, providing insights on how they correlate and on how to consider the individual or grouped behaviour to create discriminant diagrams [12,14]. Our results compare well with the literature chemical elements that better discriminate between magnetite from different metallogenic environments.

The prevalence of Ni, Co, and V as the most relevant elements in the Reference Dataset is in accordance with the results outlined in [14] in which the authors' proposed diagram includes Co and V as important elements for the discrimination of ore deposits. Ni is explained, in the case of the Alvito deposit, as an element that is possibly reflecting a stronger influence of magmatic-derived fluids from the near gabbro-diorite complex.

The PC1–PC2 plot (Figure 4) effectively discriminates most of the deposit types, with the exception of the porphyry and VMS deposits that overlap several of the other deposit types. Our results are not comparable with the ones from [14] since they do not include any sample from VMS deposits and have limited samples of magnetite with a porphyry type provenance. The authors of [12] studied magnetite from porphyry and skarn deposits and the discrimination between these two types of deposits using magnetite composition is possible, but with a bigger dispersion of data in porphyry deposits, whereas in the skarn deposits, the data are clustered. The PC1–PC2 diagram from these authors shows some degree of overlapping, similar to what was found in our case studies.

The most relevant elements for the Case Study Dataset are Zn, Co, and Al for PC1 and Mg, Mn, and Ti for PC2. These elements are also in good accordance with the conclusions presented in [12], which involves the study of skarn and porphyry Cu deposits. Sn is the exception and is considered by the authors as related to the regional geological context. In our case study, the geological context is favourable to the enrichment in Zn, which might be related to the relative enrichment in Pb-Zn of the hydrothermal systems in the study area, and a relative depletion in Sn.

Some authors [42] identified Mg and Mn as the elements that most contribute to the definition of skarn deposits, and Si, Ti, and Al as the ones that are more influenced by magmatic fluids. Zn and Co, as chalcophile elements, are influenced by the partition between sulphides and magnetite. These elements are in good correlation with the elements that define the PC1 and PC2 for our case study, clearly separating the skarn deposits (Alvito and Azenhas) from the exhalative-sedimentary-related, late-altered deposits by hydrothermal fluids (Monges, Serrinha, and Vale da Arca). The Orada deposit cluster is nearer to these late deposits, which determines that these magnetites and their deposit type must be further investigated.

5.2. ML Models in the Reference Dataset

The tested ML methods, i.e., random forest, naïve Bayes, support vector machine, multinomial linear regression, k-nearest neighbour, and artificial neural networks, all have an overall accuracy higher than 70% using the F1-score, which indicates a good prediction.

The random forest model is the one that performed better with an accuracy of 92% using the F1-score measure. This has also been referred to in [15] with a test dataset of more than 17 k observations, obtaining an accuracy of 81%. This lower accuracy compared with our results is not an indication of a better performance of our model but might be because of the immensely bigger dataset, i.e., the number of magnetite analyses used as the training dataset for the model. This necessarily lowers the accuracy by having more unconstrained observations implied in the classification of mineral deposit types. The authors recognized this constraint in their paper but consider that their number of samples might be at the peak of performance. Our results suggest that the overall accuracy obtained using the ML methods permits the use of a Train Dataset with less observations, more balanced in terms of deposit types, with exceptional results when compared with the classical diagram

approach, as also concluded in their work. The authors reported that V, Ni, and Mg are the most important elements in their model, whereas our RF model is dominated by Mg, Zn, and Ni.

The authors that tested SVM models [14] using 876 observations in magmatic and hydrothermal systems obtained F1-scores that are within this range. The average F1-score presented by these authors is 85%, whereas ours is 83%, hence revealing that our dataset and model compares well with this one. The authors [14] identified Mg, Mn, Al, Ti, V, and Co as the key factors for distinguishing their mineral deposit cases. In our case, the most important variables, using the SVM model, are Ga, Al, Co, and V, coinciding in most of the elements, with the exception of Ga that the authors did not take into account.

For discriminating IOA from IOCG deposit types, other authors [13] used an RF model with a train set of 913 LA-ICP-MS observations combined with 877 EPMA observations. Their model obtained an accuracy of 91%, which compares well with our F1-score of 92% using the same machine learning model. These authors identified Si, Mg, and V as the most important elements in their classification, and our RF model identifies Mg, Zn, and Ni as the most important ones. This might be due to the higher number of deposit types considered in our model.

Considering the selected chemical elements, the Reference Dataset demonstrates a difficulty in identifying VMS deposit types, which is in agreement with other authors' conclusions [15]. The ML model was the one that performed better in classifying VMS deposits, showing an MLR model with 86% true positive values.

Table 3 summarizes the results obtained and compares them with bibliographic data.

Table 3. Comparison of this study's results with bibliographic data. The accuracy values in our study are measured as F1-score.

ML Model	[13]	[14]	[15]	This Study
RF	91% Si, Mg, V		81% V, Ni, Mg	92% Mg, Zn, Ni
NB			58% Not defined	89% Ni, Ti, V
SVM		85% Mg, Mn, Al, Ti, V, Co		83% Ga, Al, Co, V
MLR				88% Co, Mg, Ga, Ni
KNN			69% Not defined	77% Mg, Al, Ti
ANN				76% Mg, Mn, V

This analysis allows us to consider the RF as the best model for identifying the type of deposit, followed by NB and MLR models.

5.3. ML Applied to the Case Study Dataset

For the Case Study Dataset, the scenery is somewhat different. The RF model has a lower mean assertiveness (69%), and the MLR has the highest (89%), but one must remember that this is not a measure of the accuracy of the model but an indicator of how the model maintains its classification in the same group.

With the RF model, Azenhas, Serrinha, and Vale da Arca are the deposits that have a higher assertiveness (>50%). Azenhas and Alvito are classified as skarns, while Serrinha and Vale da Arca are classified as BIF. Monges classification is at a stalemate between BIF and VMS. Considering the proposed classification from published works [5,6,17,56], the classification of Alvito and Azenhas deposits, using magnetite composition, as skarn deposits is correct. For Monges, Serrinha, and Vale da Arca, their attribution to BIF as well as to VMS might also be considered as correct, as their exhalative-sedimentary provenance combined with late hydrothermal alterations results in a variable trace element composition

in these magnetites. Moreover, the indecision for the Alvito deposit between skarn and porphyry (i.e., with magmatic fluid input) is also an indication of the probable mixed nature of this deposit with a strong influence from the nearby Gabro-Dioritic suite. Orada deposit classification is very disperse, with most votes falling in the VMS deposit type, also evidencing a somewhat mixed nature of this deposit.

The NB model has a somewhat similar result as RF. However, for the Azenhas deposit, it falls clearly within the porphyry field. The classification of the Orada deposit in this model is clearly in the VMS type.

SVM and MLR models provide similar identification results, clearly considering Alvito and Azenhas as skarns, whereas Monges, Serrinha, and Vale da Arca are in the porphyry deposit type. Both models classify Orada as a BIF deposit. Hence, for Alvito and Azenhas, there is a clear agreement as to the skarn classification and there is an alternative classification of Monges, Serrinha, and Vale da Arca as porphyry deposit types. This might be due to the fact that these models attribute a higher importance to elements such as Co and Ga that are usually found in a higher concentration in magnetite from porphyry deposits when compared to magnetite from BIF deposits (c.f. Table 1). The KNN model is similar to the SVM and MLR models; however, it classifies Serrinha as a skarn deposit, possibly due to the higher Mg value, which is the most important element in this model and is characteristic of skarn deposits as a result of metasomatic alteration.

The ANN model provides significantly different classifications that are hard to match with any interpretation of the models. Alvito and Orada are classified as being a skarn deposit and Monges is classified as a porphyry deposit. Azenhas is classified as a BIF deposit, and finally Serrinha and Vale da Arca are classified as VMS. This is an indicator that the results from this model must be discarded or tested with better-tuned parameters.

It must be noted that in the diagram [7] (Figure 3), all the case study deposits fall in the skarn deposit type and most of the used models are able to distinguish and better identify the deposit type for these cases.

6. Conclusions

Meaningful methodological and metallogenic conclusions can be drawn from the results obtained. These are summarized as the following:

- The use of a limited dataset of trace element geochemistry from magnetites combined with most of the machine learning models has demonstrated to be a reliable tool to provide accurate classifications of mineral deposit types.
- The different machine learning methods tested provide accurate classifications of the mineral deposit types; however, random forest, naïve Bayes, and multinomial logistic regression are identified as the most accurate. Artificial neural networks have the worst performance.
- Combining all the proposed classifications, Alvito and Azenhas are confidently classified as skarns. Monges, Serrinha, and Vale da Arca are either classified as porphyry or BIF. Orada is cryptically classified either as BIF or VMS.
- The application of machine learning models to relevant case studies provides meaningful results that can be used in the interpretation of not only the mineral deposit types but also offered clues to identify mixed or more complex origins.
- The application of these methods to exploration projects is a powerful tool that sheds new light on magnetite interpretation that might provide clues for new mineral exploration targets.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/min13081009/s1>, Table S1: Confusion matrix for the RF model with 500 trees and 70% of data as Train and 30% as Test; Table S2: Confusion matrix for the NB model; Table S3: Confusion matrix for the SVM model; Table S4: Confusion matrix for the MLR model; Table S5: Confusion matrix for the KNN model; Table S6: Confusion matrix for the ANN model; Table S7: Cross reference table for deposit types using RF model; Table S8: Cross reference

table for deposit types using NB model; Table S9: Cross reference table for deposit types using SVM model; Table S10: Cross reference table for deposit types using MLR model; Table S11: Cross reference table for deposit types using KNN model; Table S12: Cross reference table for deposit types using ANN model.

Author Contributions: Conceptualization, P.N. and M.M.; methodology, M.M. and P.N.; validation, P.N. and M.M.; investigation, M.M. and P.N.; data curation, M.M.; writing—original draft preparation, P.N. and M.M.; writing—review and editing, P.N. and M.M.; visualization, M.M.; supervision, P.N.; project administration, M.M. and P.N.; funding acquisition, M.M. and P.N. All authors have read and agreed to the published version of the manuscript.

Funding: M. Maia acknowledges the financial support of Fundação para a Ciência e Tecnologia (FCT; Portuguese Science and Technology Foundation) through the PhD grant SFRH/BD/145049/2019, as well as the financial support provided by the Society of Economic Geologists Foundation (SEGF) through the Student Research Grant–Hugh McKinstry Fund.

Data Availability Statement: All the data is published in Supplementary materials or in the referenced articles.

Acknowledgments: Authors acknowledge the funding provided by the Institute of Earth Sciences (ICT) through the COMPETE 2020 (UIDB/04683/2020 and UIDP/04683/2020) under the reference POCI-01-0145-FEDER-007690. This work is also a contribution to the project “ZOM-3D Metallogenic Modelling of Ossa-Morena Zone: Valorisation of the Alentejo Mineral Resources” (ALT20-03-0145-FEDER-000028), funded by Alentejo 2020 (Regional Operational Program of Alentejo) through the FEDER/FSE/FEEL.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, W.; Ying, Y.-C.; Bai, T.; Zhang, J.-J.; Jiang, S.-Y.; Zhao, K.-D.; Shin, D.; Kynicky, J. In situ major and trace element analysis of magnetite from carbonatite related complexes: Implications for petrogenesis and ore genesis. *Ore Geol. Rev.* **2019**, *107*, 30–40. [[CrossRef](#)]
2. Liu, Y.; Fan, Y.; Zhou, T.; Xiao, X.; White, N.C.; Thompson, J.; Hong, H.; Zhang, L. Geochemical characteristics of magnetite in Longqiao skarn iron deposit in the Middle-Lower Yangtze Metallogenic Belt, Eastern China. *Miner. Depos.* **2019**, *54*, 1229–1242. [[CrossRef](#)]
3. Ayupova, N.R.; Novoselov, K.A.; Maslennikov, V.; Melekestseva, I.Y.; Hollis, S.P.; Artemyev, D.A.; Tessalina, S.G. The formation of magnetite ores of the Glubochenskoe deposit, Turgai iron belt, Russia: New structural, mineralogical, geochemical, and isotopic constraints. *Miner. Depos.* **2020**, *56*, 103–123. [[CrossRef](#)]
4. Qi, Y.; Hu, R.; Gao, J.; Gao, W.; Gong, H. Trace element characteristics of magnetite: Constraints on the genesis of the Lengshuikeng Ag–Pb–Zn deposit, China. *Ore Geol. Rev.* **2021**, *129*, 103943. [[CrossRef](#)]
5. Maia, M.; Barrulas, P.; Nogueira, P.; Mirão, J.; Noronha, F. In situ LA-ICP-MS trace element analysis of magnetite as a vector towards mineral exploration: A comparative case study of Fe-skarn deposits from SW Iberia (Ossa-Morena Zone). *J. Geochem. Explor.* **2022**, *234*, 106941. [[CrossRef](#)]
6. Maia, M.; Barrulas, P.; Nogueira, P.; Mirão, J.; Noronha, F. Combining $\delta^{18}\text{O}$ isotope data and in-situ LA-ICP-MS trace element analysis of magnetite as a proxy for ore genesis: Constraints on the formation of Fe deposits from Ossa-Morena Zone (SW Iberian Peninsula). *J. Geochem. Explor.* **2023**, *245*, 107140. [[CrossRef](#)]
7. Dupuis, C.; Beaudoin, G. Discriminant diagrams for iron oxide trace element fingerprinting of mineral deposit types. *Miner. Depos.* **2011**, *46*, 319–335. [[CrossRef](#)]
8. Dare, S.A.S.; Barnes, S.-J.; Beaudoin, G.; Méric, J.; Boutroy, E.; Potvin-Doucet, C. Trace elements in magnetite as petrogenetic indicators. *Miner. Depos.* **2014**, *49*, 785–796. [[CrossRef](#)]
9. Nadoll, P.; Angerer, T.; Mauk, J.L.; French, D.; Walshe, J. The chemistry of hydrothermal magnetite: A review. *Ore Geol. Rev.* **2013**, *61*, 1–32. [[CrossRef](#)]
10. Nadoll, P.; Mauk, J.L.; Leveille, R.A.; Koenig, A.E. Geochemistry of magnetite from porphyry Cu and skarn deposits in the southwestern United States. *Miner. Depos.* **2015**, *50*, 493–515. [[CrossRef](#)]
11. Makvandi, S.; Ghasemzadeh-Barvarz, M.; Beaudoin, G.; Grunsky, E.C.; McClenaghan, M.B.; Duchesne, C. Principal component analysis of magnetite composition from volcanogenic massive sulfide deposits: Case studies from the Izok Lake (Nunavut, Canada) and Halfmile Lake (New Brunswick, Canada) deposits. *Ore Geol. Rev.* **2016**, *72*, 60–85. [[CrossRef](#)]
12. Canil, D.; Grondahl, C.; Lacourse, T.; Pisiak, L.K. Trace elements in magnetite from porphyry Cu–Mo–Au deposits in British Columbia, Canada. *Ore Geol. Rev.* **2016**, *72*, 1116–1128. [[CrossRef](#)]
13. Hong, S.; Zuo, R.; Juang, X.; Xiong, Y. Distinguishing IOCG and IOA deposits via random forest algorithm based on magnetite composition. *J. Geochem. Explor.* **2021**, *230*, 106859. [[CrossRef](#)]

14. Hu, B.; Zeng, L.-P.; Liao, W.; Wen, G.; Hu, H.; Li, M.Y.H.; Zhao, X. The Origin and Discrimination of High-Ti Magnetite in Magmatic-Hydrothermal Systems: Insight from Machine Learning Analysis. *Econ. Geol.* **2022**, *117*, 1613–1627. [[CrossRef](#)]
15. Bédard, É.; De Vazelhes, V.D.; Beaudoin, G. Performance of predictive supervised classification models of trace elements in magnetite for mineral exploration. *J. Geochem. Explor.* **2022**, *236*, 106959. [[CrossRef](#)]
16. Gregory, D.D.; Cracknell, M.J.; Large, R.R.; McGoldrick, P.; Kunh, S.; Maslennikov, V.V.; Baker, M.J.; Fox, N.; Belousov, I.; Figueroa, M.C.; et al. Distinguishing Ore Deposit Type and Barren Sedimentary Pyrite Using Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry Trace Element Data and Statistical Analysis of Large Data Sets. *Econ. Geol.* **2019**, *114*, 771–786. [[CrossRef](#)]
17. Mateus, A.; Munhá, J.; Inverno, C.; Matos, J.X.; Martins, L.; Oliveira, D.; Jesus, A.; Salgueiro, R. Mineralizações no sector português da Zona de Ossa-Morena. In *Geologia de Portugal*; Dias, R., Araújo, A., Terrinha, P., Kullberg, J.C., Eds.; Escolar Editora: Lisboa, Portugal, 2013; Volume 1, pp. 577–619.
18. Tornos, F.; Inverno, C.M.C.; Casquet, C.; Mateus, A.; Ortiz, G.; Oliveira, V. The metallogenic evolution of the Ossa-Morena Zone. *J. Iber. Geol.* **2004**, *30*, 143–181.
19. Ribeiro, A.; Munhá, J.; Dias, R.; Mateus, A.; Pereira, E.; Ribeiro, L.; Fonseca, P.; Araújo, A.; Oliveira, O.; Romão, J.; et al. Geodynamic evolution of the SW Europe Variscides. *Tectonics* **2007**, *26*, TC6009. [[CrossRef](#)]
20. Chichorro, M. A Evolução Tectónica da Zona de Cisalhamento de Montemor-o-Novo (Sudoeste da Zona de Ossa-Morena—Área de Santiago do Escoural—Cabrela). Ph.D. Thesis, University of Évora, Évora, Portugal, 2016; p. 569.
21. Pereira, M.F.; Silva, J.B.; Chichorro, M.; Moita, P.; Santos, J.F.; Apraiz, A.; Ribeiro, C. Crustal growth and deformational processes in the northern Gondwana margin: Constraints from the Évora Massif (Ossa-Morena zone, southwest Iberia, Portugal). In *Special Paper 423: The Evolution of the Rheic Ocean: From Avalonian-Cadomian Active Margin to Alleghenian-Variscan Collision*; The Geological Society of America: Boulder, CO, USA, 2007; pp. 333–358. [[CrossRef](#)]
22. Salgueiro, R. Caracterização e Génese das Mineralizações de Magnetite—Sulfuretos em Monges (Santiago do Escoural, Montemor-o-Novo) e Ensaio Comparativo com as Suas Congéneres em Orada-Vale de Pães (Serpa-Vidigueira). Ph.D. Thesis, University of Lisbon, Lisbon, Portugal, 2011; 524p.
23. Julivert, M.; Fontboté, J.M.; Ribeiro, A.; Conde, L. *Mapa Tectónico de La Península Ibérica y Baleares (Tectonic Map of the Iberian Peninsula and Balearian Islands)*; IGME-SPI, Instituto Geológico y Minero de España: Madrid, Spain, 1972; p. 113.
24. Jesus, A.P.; Mateus, A.; Benoit, M.; Tassinari, C.C.G.; Bento dos Santos, T. The timing of sulfide segregation in a Variscan synorogenic gabbroic layered intrusion (Beja, Portugal): Implications for Ni-Cu-PGE exploration in orogenic settings. *Ore Geol. Rev.* **2020**, *126*, 103767. [[CrossRef](#)]
25. Andrade, A.; Silva, J.M.; Arruda, C.R.; Gameiro, J.C.S. Minas de Ferro de Montemor-o-Novo. *Serviço De Fom. Min.* **1949**, *15*, 125.
26. Goinhas, A.C.; Martins, L.M.P. Área metalífera de Montemor-o-Novo—Casa Branca (Baixo Alentejo, Portugal). *Estud. Notas E Trab.* **1986**, *28*, 119–148.
27. Carvalhosa, A.B.; Zbyszewski, G. *Carta Geológica de Portugal à Escala 1: 50,000: Folha 40-C Viana do Alentejo*; Direção Geral de Minas e Serviços Geológicos: Lisboa, Portugal, 1971.
28. Gomes, E.M.C. Metamorfismo de rochas carbonatadas siliciosas da região de Alvito (Alentejo, Sul de Portugal). Ph.D. Thesis, University of Coimbra, Coimbra, Portugal, 2000; 240p.
29. Gomes, E.M.C.; Fonseca, P.E. Eventos metamórfico/metassomáticos tardi-variscos na região de Alvito (Alentejo, sul de Portugal). In *Cadernos Xeológicos de Laxe*; Instituto Universitario de Xeoloxía Isidro Parga Pondal: A Coruña, Spain, 2006; Volume 31, pp. 67–85, ISSN 0213-4497.
30. *Carta Geológica da Região de Pedrógão-Orada à escala 1:10,000*; Serviços de Fomento Mineiro: Lisboa, Portugal, 1965.
31. Pin, C.; Fonseca, P.E.; Paquette, J.-L.; Castro, P.; Matte, P. The ca. 350 Ma Beja Igneous Complex: A record of transcurrent slab break-off in the Southern Iberia Variscan Belt? *Tectonophysics* **2008**, *461*, 356–377. [[CrossRef](#)]
32. Araújo, A.; Fonseca, P.; Munhá, J.; Moita, P.; Pedro, J.; Ribeiro, A. The Moura Phyllonitic Complex: An Accretionary Complex related with obduction in the Southern Iberia Variscan Suture. *Geodin. Acta* **2005**, *18*, 375–388. [[CrossRef](#)]
33. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, caret1–caret26. [[CrossRef](#)]
34. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
35. Majka, M. *NaiveBayes: High Performance Implementation of the Naive Bayes Algorithm in R, R Package Version 0.9.7 2019*; CRAN: Online repository, 2020.
36. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0.
37. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group, Formerly: E1071, R Package Version 1.7-13*; TU Wien: Vienna, Austria, 2023.
38. Araújo, J.C.S.; Lobato, L.M. Depositional model for banded iron formation host to gold in the Archean Rio das Velhas greenstone belt, Brazil, based on geochemistry and LA-ICPMS magnetite analyses. *J. S. Am. Earth Sci.* **2019**, *97*, 102205. [[CrossRef](#)]
39. Wang, Y.-J.; Zhu, W.-G.; Huang, H.-Q.; Bai, Z.-J.; Zhong, H.; Yao, J.-H.; Fan, H.-P. Geochemistry of magnetite from the giant Paleoproterozoic Dahongshan Fe-Cu deposit, SW China: Constraints on nature of ore-forming fluids and depositional setting. *Ore Geol. Rev.* **2020**, *118*, 103361. [[CrossRef](#)]
40. Huang, X.-W.; Sappin, A.-A.; Boutroy, É.; Beaudoin, G.; Makvandi, S. Trace element composition of igneous and hydrothermal magnetite from porphyry deposits: Relationship to deposit subtypes and magmatic affinity. *Econ. Geol.* **2019**, *114*, 917–952. [[CrossRef](#)]

41. La Cruz, N.L.; Ovalle, J.T.; Simon, A.C.; Konecke, B.A.; Barra, F.; Reich, M.; Leisen, M.; Childress, T.M. The Geochemistry of Magnetite and Apatite from the El Laco Iron Oxide–Apatite Deposit, Chile: Implications for Ore Genesis. *Econ. Geol.* **2020**, *115*, 1461–1491. [[CrossRef](#)]
42. Hu, X.; Chen, H.; Zhao, L.; Han, J.; Xia, X. Magnetite geochemistry of the Longqiao and Tieshan Fe–(Cu) deposits in the Middle-Lower Yangtze River Belt: Implications for deposit type and ore genesis. *Ore Geol. Rev.* **2017**, *89*, 822–835. [[CrossRef](#)]
43. Makvandi, S. Indicator Mineral Exploration Methodologies for VMS Deposits Using Geochemistry and Physical Characteristics of Magnetite. Ph.D. Thesis, Université Laval, Quebec City, QC, Canada, 2015; 235p.
44. Makvandi, S.; Ghasemzadeh-Barvarz, M.; Beaudoin, G.; Grunsky, E.C.; McClenaghan, M.B.; Duchesne, C.; Boutroy, E.C. Partial least squares-discriminant analysis of trace element compositions of magnetite from various VMS deposit subtypes: Application to mineral exploration. *Ore Geol. Rev.* **2016**, *78*, 388–408. [[CrossRef](#)]
45. Palarea-Albaladejo, J.; Martín-Fernández, J.A. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* **2015**, *143*, 85–96. [[CrossRef](#)]
46. Hron, K.; Templ, M.; Filzmoser, P. Imputation of missing values for compositional data using classical and robust methods. *Comput. Stat. Data Anal.* **2010**, *54*, 3095–3107. [[CrossRef](#)]
47. Canil, D.; Lacourse, T. Geothermometry using minor and trace elements in igneous and hydrothermal magnetite. *Chem. Geol.* **2020**, *541*, 119576. [[CrossRef](#)]
48. O'Brien, J.J.; Spry, P.G.; Nettleton, D.; Xu, R.; Teale, G.S. Using Random Forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb–Zn–Ag deposits in the Broken Hill domain, Australia. *J. Geochem. Explor.* **2015**, *149*, 74–86. [[CrossRef](#)]
49. Zhong, R.; Deng, Y.; Li, W.; Nanyushevsky, L.V.; Cracknell, M.J.; Belousov, I.; Chen, Y.; Li, L. Revealing the multi-stage ore-forming history of a mineral deposit using pyrite geochemistry and machine learning-based data interpretation. *Ore Geol. Rev.* **2021**, *133*, 104079. [[CrossRef](#)]
50. Ibrahim, A.M.; Bennett, B.; Isiaka, F. The optimisation of Bayesian classifier in predictive spatial modelling for secondary mineral deposits. *Procedia Comput. Sci.* **2015**, *61*, 478–485. [[CrossRef](#)]
51. Mohammadi, N.M.; Hezarkhani, A. Application of support vector machine for the separation of mineralised zones in the Takht-e-Gonbad porphyry deposit, SE Iran. *J. Afr. Earth Sci.* **2018**, *143*, 301–308. [[CrossRef](#)]
52. Maepa, F.; Smith, R.S.; Tessema, A. Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geol. Rev.* **2021**, *130*, 103968. [[CrossRef](#)]
53. Kaplan, U.E.; Topal, E. A new ore grade estimation using combine machine learning algorithms. *Minerals* **2020**, *10*, 847. [[CrossRef](#)]
54. Darwish, Y.Z.; Embaby, A.K.; Sharafeldin, H.E.; Farag, H.A.; El Kholy, D.M.; Selim, S.M. Developing a Forecasting model for uranium occurrence in GII, Northeastern Desert, Egypt using artificial neural networks. *J. Radiat. Res. Appl. Sci.* **2022**, *15*, 100468. [[CrossRef](#)]
55. Gevrey, M.; Dimopoulos, I.; Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* **2003**, *160*, 249–264. [[CrossRef](#)]
56. Salgueiro, R.; Mateus, A.; Inverno, C. Mineralizações de magnetite e sulfuretos de monges (Santiago do Escoural, Montemor-o-Novo), Vale de Pães (Cuba-Vidigueira) e Orada (Pedrógão, Serpa): Síntese de ensaio comparativo. In *Boletim de Minas*; LNEG: Lisboa, Portugal, 2012; Volume 47, pp. 27–30.

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.