

## Article

# Machine Learning in Discriminating Active Volcanoes of the Hellenic Volcanic Arc

Athanasios G. Ouzounis \*  and George A. Papakostas 

HUMAIN-Lab, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece; gpapak@cs.ihu.gr

\* Correspondence: athouzo@ihu.gr; Tel.: +30-6945260121

**Abstract:** Identifying the provenance of volcanic rocks can be essential for improving geological maps in the field of geology and providing a tool for the geochemical fingerprinting of ancient artifacts like millstones and anchors in the field of geoarchaeology. This study examines a new approach to this problem by using machine learning algorithms (MLAs). In order to discriminate the four active volcanic regions of the Hellenic Volcanic Arc (HVA) in Southern Greece, MLAs were trained with geochemical data of major elements, acquired from the GEOROC database, of the volcanic rocks of the Hellenic Volcanic Arc (HVA). Ten MLAs were trained with six variations of the same dataset of volcanic rock samples originating from the HVA. The experiments revealed that the Extreme Gradient Boost model achieved the best performance, reaching 93.07% accuracy. The model developed in the framework of this research was used to implement a cloud-based application which is publicly accessible at This application can be used to predict the provenance of a volcanic rock sample, within the area of the HVA, based on its geochemical composition, easily obtained by using the X-ray fluorescence (XRF) technique.

**Keywords:** machine learning; Hellenic Volcanic Arc; geochemistry; geoarchaeology; geochemical fingerprinting



**Citation:** Ouzounis, A.G.; Papakostas, G.A. Machine Learning in Discriminating Active Volcanoes of the Hellenic Volcanic Arc. *Appl. Sci.* **2021**, *11*, 8318. <https://doi.org/10.3390/app11188318>

Academic Editor: Flavio Cannavò

Received: 30 June 2021

Accepted: 5 September 2021

Published: 8 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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

The Hellenic Volcanic Arc (HVA) in Southern Greece consists of four volcanic regions (Figure 1), namely the Methana region, which includes Methana, Aigina, and Poros in the west; the Melos region, consisting of Melos, Antimelos, and Kimolos in the southwest; the Santorini region, consisting of Santorini, Kolombus, and Christianoi in the southeast; and the Nisyros region consisting of Nisyros, Giali, and Pergousa in the east. All active volcanoes of the HVA were formed by a converging tectonic setting, producing Island Arc Basalts (IABs) belonging to the calc-alkaline and high-K calc-alkaline series. The first activity took place during the Pliocene (5 m.y.) and continues to the present day [1–3]. The extent of the HVA is well-mapped on the surface [1], but little is known about its extent under the seafloor. A large part of the HVA is buried under the sediments of the Southern Aegean Sea. Future deep-sea drilling for mineral prospecting will provide rock samples to produce more precise geological borders for the HVA. In order for these rock samples to be used for mapping, they have to be correlated to one of the four active volcanic regions. Furthermore, volcanic rocks from the HVA have been used to produce artifacts like millstones and anchors from prehistoric times. These artifacts can be found scattered all around the Mediterranean. One of the questions about these artifacts in archaeology is the source of the original material. By answering that question, ancient trade routes can be revealed.

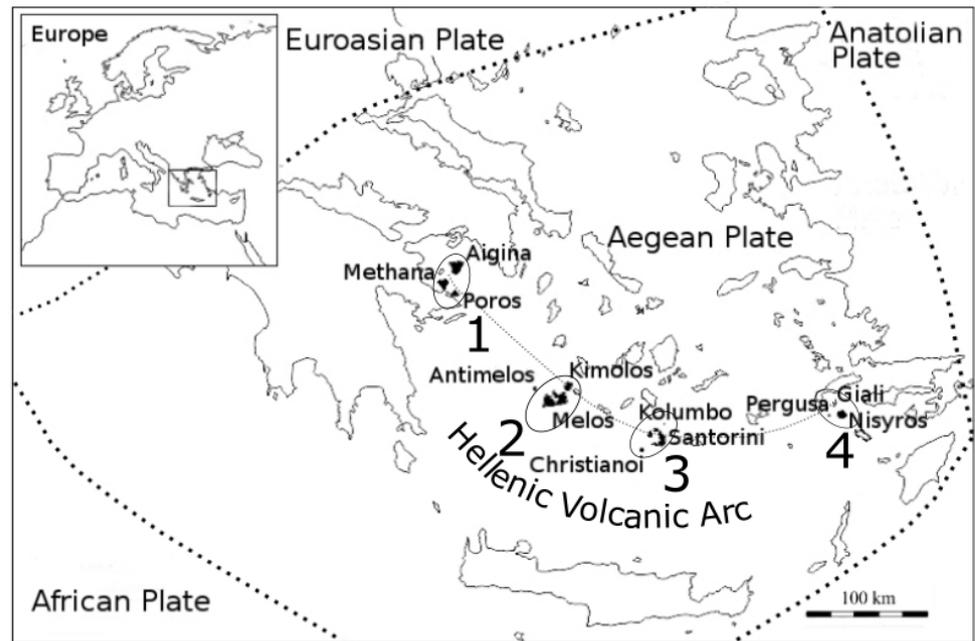


Figure 1. The Hellenic Volcanic Arc: 1. Methana group, 2. Melos group, 3. Santorini group, and 4. Nisyros group.

One obvious approach used to correlate volcanic rocks and fingerprint ancient artifacts is to use the geochemistry of these rocks to identify their provenance. By studying the binary  $K_2O$  versus  $SiO_2$  diagram [4] of the HVA compiled from 1500 data samples (Figure 2) [2], major overlaps of the data points can be seen. It is obvious that no discrimination between the volcanic regions of the HVA can be established based on this diagram.

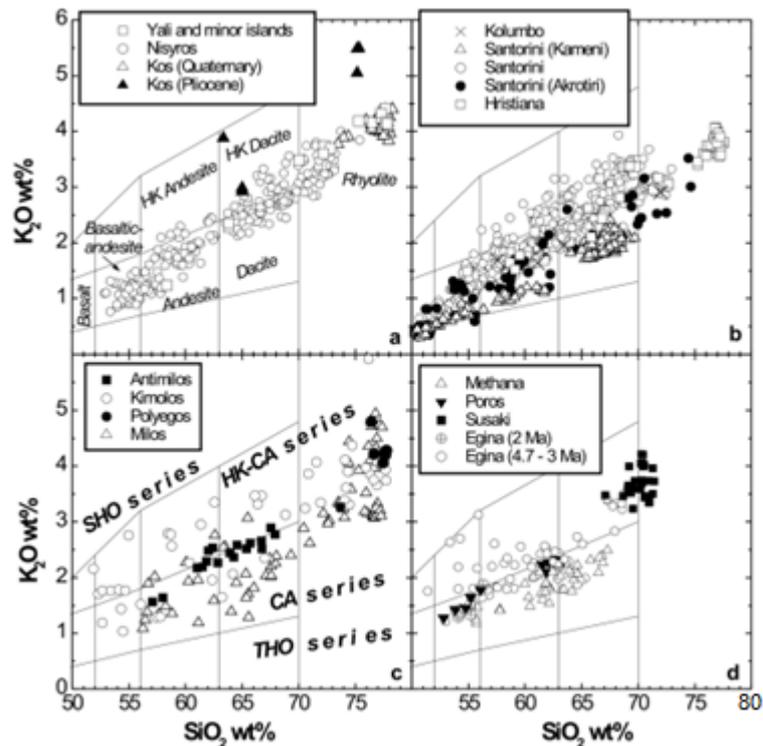


Figure 2.  $K_2O$  versus silica classification diagrams [4] for (a) Nisyros group (b) Santorini group (c) Melos group and (d) Methana group.

The diagram in Figure 2 is based on binary and ternary diagrams, known as discrimination diagrams, which were introduced in the 1970s to discriminate geotectonic settings based on the trace element [5] and major element [6] geochemical data of basalts and sandstones, respectively. The main geotectonic settings are the boundaries of either converging or diverging lithospheric plates where volcanic activity is observed, resulting in the formation of basalts, the most common rock formed by magma pouring out from volcanoes originating from the Earth's mantle. Basalts can also be formed at hot spots far away from tectonic boundaries, which are known as Ocean Island Basalts (OIB). Given the small amount of data available and the limited areas under examination at the time, discrimination diagrams performed well and this is why they are still used and are continuously being evolved [7–9]. Since 1970, however, the amount of rock data acquired from all over the world has rapidly increased. Huge datasets containing information on the chemical analysis of rock samples have become openly available from databases such as GEOROCK [10] and PetDB [11]. These datasets have been used to prove that conventional discrimination diagrams are inaccurate in many cases [12,13] because of the overlap of the sample points, as is the case in the diagram in Figure 2.

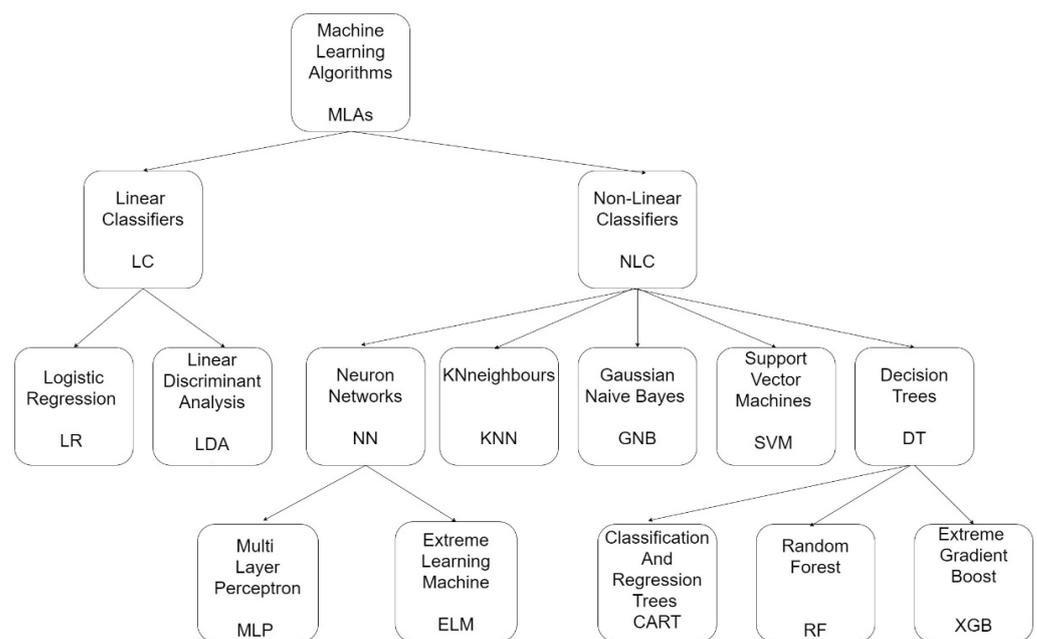
Machine learning (ML) has been used in many fields of science to study the hidden patterns in large datasets [14]. After 2000, attempts have been made to use this technology to discriminate tectonic settings. The first successful attempt to discriminate tectonic settings using ML was made using Decision Trees (DT) in 2006 [15]. Ten years later, a Support Vector Machine (SVM) was tested on a dataset using the geochemical and isotopic data of basalts to discriminate tectonic settings [16]. One year later, SVM, Random Forest (RF), and Sparse Multinomial Regression (SMR) were compared using 20 elements and five isotopic ratios of basaltic rocks. The accuracy achieved was over 83% [17]. In 2019, chemical data derived from olivine, a major mineral found in basalts, were used to compare the accuracy of Logistic Regression (LR), Naïve Bayes (NB), RF, and Multi-Layer Perceptron (MLP) for discriminating tectonic settings [18]. The same group introduced a new method named SONFIS incorporating fuzzy logic [19]. The latest published work used the Logistic Regression Analysis (LRA) algorithm with the geochemical data of monazite, a mineral found in basalt rock [20]. In 2017, SVM was used to discriminate volcanic centers in Italy [21].

In this research, the aforementioned methodology used for discriminating tectonic settings was applied in order to evaluate its suitability for distinguishing volcanic regions belonging to one tectonic setting. This was accomplished by comparing the performance of 10 MLAs using six variations of the initial dataset retrieved from the GEOROCK databases. Only major elements were used, as usable trace element data were only available for 10% of the samples of the dataset. Finally a cloud-based application (CBA) was implemented using the best performing model with the ability to correlate a volcanic rock sample from the HVA with one of the four volcanic regions of the South Aegean, based on the geochemical composition of 10 major elements ( $\text{SiO}_2$ ,  $\text{TiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3\text{T}$ ,  $\text{CaO}$ ,  $\text{MgO}$ ,  $\text{MnO}$ ,  $\text{K}_2\text{O}$ ,  $\text{Na}_2\text{O}$ , and  $\text{P}_2\text{O}_5$ ). This application is publicly accessible at [www.geology.gr](http://www.geology.gr) (accessed on 6 September 2021).

## 2. Materials and Methods

### 2.1. Machine Learning Algorithms

The goal of ML is to construct a model for making predictions [22]. There are two essential parts to constructing a model: (1) the Machine Learning Algorithm (MLA) [23] and (2) the data. A wide variety of MLAs have been developed since the 1960s, and their improvement has never ceased. The choice of the best-performing algorithm is determined heuristically for a given problem. The best learning algorithm is considered to be the one that has low bias and accurately relates the input to the output values, achieving low variance, meaning that it can successfully predict the same result using different datasets. The latter is also called generalization. Ten MLAs, both linear and non-linear, were used in this study (Figure 3).



**Figure 3.** Outline of the Machine Learning Algorithms used.

Of the linear algorithms, Logistic Regression (LR) [24] and Linear Discriminant Analysis (LDA) [25] were used. The non-linear algorithms tested were k-Nearest Neighbor (kNN) [26], Gaussian Naive Bayes (GNB) [27], and Support Vector Machine (SVM) [28]. Furthermore, Multi-Layer Perceptron (MLP) [29,30] and Extreme Learning Machine (ELM) [31,32] belonging to the Neural Network Group were evaluated. Finally, decision trees were tested, namely Classification and Regression Trees (CART) [33], Random Forest (RF) [34–36], and Extreme Gradient Boost (XGB) [37,38].

#### 2.1.1. Logistic Regression

The LR algorithm [24] is an extension of the linear regression algorithm [39], where the goal is to find the ‘line of best fit’ in order to map the relationship between the dependent variable and the independent variables. The LR algorithm predicts whether something belongs to a class or not, i.e., if it is true or false. LR is mainly used for binary classification but can be extended for multiclass classification. In our case, Multinomial Logistic Regression was used.

#### 2.1.2. Linear Discriminant Analysis

LDA is a supervised algorithm which reduces dimensionality. LDA is a two-class classifier but can be generalized for multi-class problems. The linear discriminants are the axes, which maximize the distances of multiple classes [25].

#### 2.1.3. k-Nearest Neighbor

In the kNN classifier [26], the distances of the sample points are calculated from a random central point, and each data point is grouped according to this distance. It is regarded as one of the simplest MLAs. The k parameter indicates the number of points that will be taken into account. Larger values create high bias and should be avoided. Another parameter to be selected is the way the distance is computed. Some options are: (1) Euclidean distance, (2) Manhattan distance, and (3) Minkowski distance.

#### 2.1.4. Gaussian Naïve Bayes

The GNB is a probabilistic non-parametric classifier and, as the name implies, is based on Bayes’ theorem [27].

### 2.1.5. Support Vector Machine

SVM is a supervised classification technique and was initially developed for binary classification problems. The main idea is that input vectors are mapped in a non-linear way to a very high-dimension feature space. A linear decision surface is constructed in this feature space. When a dataset consisting of three features is trained by an SVM, each point is plotted in the three-dimensional space. The algorithm's goal is to classify the points by applying three planes. In the case of  $n$  features, the points are projected in an  $n$ -dimensional space, which cannot be visualized. The algorithm's goal is to classify the objects by dividing the  $n$ -dimensional space with the use of what is known as hyperplanes. Predictions are achieved by projecting new points in these divisions [28].

### 2.1.6. Multi-Layer Perceptron

An MLP is a type of feedforward Artificial Neural Network (ANN). ANNs are based on the assumption that the way the human brain operates can be mimicked artificially by creating artificial neurons named perceptrons [30]. This is achieved by providing input and output data to the algorithm, which tries, by changing the weights of its neurons, to create a function that successfully maps the relationship between the input and the output data. In the case of the MLP, at least three layers are required: the input layer, the hidden layer, and the output layer. Each node is a neuron that uses a non-linear activation function. In this way, the MLP can classify separable non-linear data [29].

### 2.1.7. Extreme Learning Machine

ELMs are essentially neural networks with only one hidden layer. According to the original authors [32], this algorithm performs faster than MLPs with good generalization. In an ELM, the hidden nodes' learning parameters, such as the input weight and biases, are randomly assigned and, according to the author, no tuning is needed. Furthermore, the output weights are determined by a simple inverse operation [31]. The only parameter tuned is the number of neurons in the hidden layer.

### 2.1.8. Classification and Regression Trees

When CART is used, classification of the samples is achieved by the binary partitioning of a feature [33]. It is not only the ease and robustness of construction that make CART a very attractive classifier. The big advantage is the possibility of interpreting the classification, which is either difficult or not possible at all with the aforementioned classifiers. Another advantage is the ability of classification trees to handle missing data. The disadvantage, however, is the overfitting problem. The tree created works well with the data used to create it but does not perform that well with new data. This is also called low bias–high variance. No parameters are set in this algorithm.

### 2.1.9. Random Forest

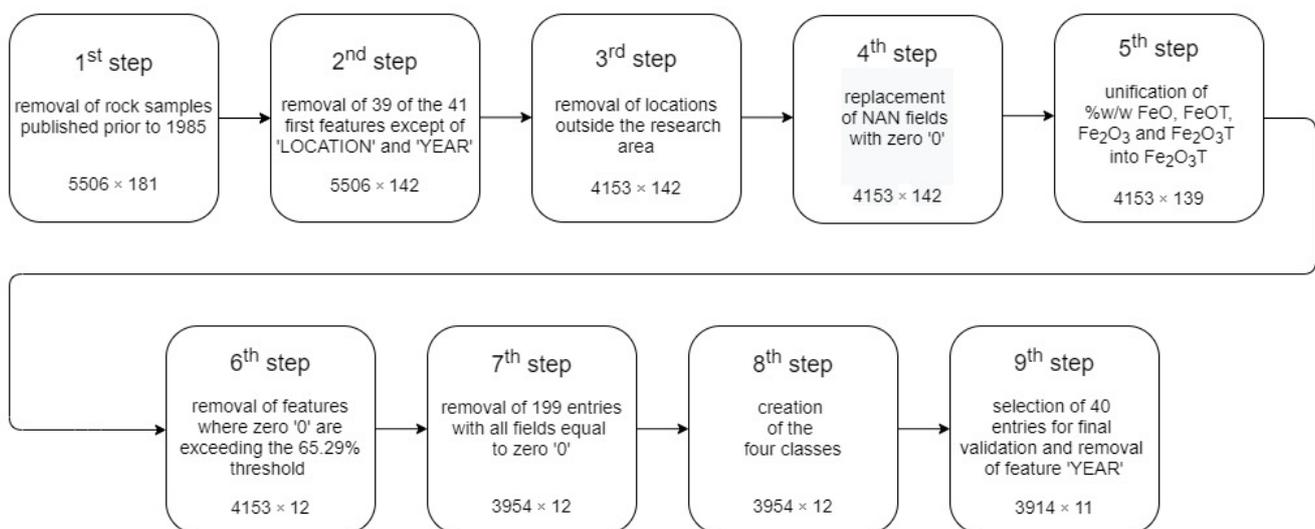
To deal with the overfitting problem of CART, the ensemble classifier RF was developed [35,36]. The main idea is to use many trees instead of one. The trees are built by using randomly selected subsets of the original data. The only constraint is that all the vectors must have the same dimensions. This is called bootstrapping. The results obtained from each tree are then put to a vote. The aggregate is the final result. This technique is known as bagging (from bootstrap aggregation) [34]. The split at each node is made by randomly selecting a subset of features. As the number of trees grows, better generalization is achieved. The parameters that can be set are: (1) the number of trees, (2) the number of features used to make the split at each node, and (3) the function used to measure the effectiveness of the split.

### 2.1.10. Extreme Gradient Boost

XGB is the latest classifier in the evolution of tree-based classifiers. It introduced tree gradient boosting [38]. Furthermore, XGB can handle sparse data [37]. According to the authors, XGB runs 10 times faster than other solutions.

### 2.2. Data Preparation

A collection of volcanic rock geochemical data, covering the active volcanic regions (4) of the HVA and most non-active pre-Pliocene volcanic regions (32) of Greece, was retrieved from the GEOROC database [10]. The initial dataset provided consists of 6229 entries containing 181 features. The first 41 features are descriptive and cover general information about the rock samples. Most notable is the year of publication of the rock sample data and the sampling location. Initially, data reduction was performed using the Python libraries Pandas [40] and Numpy [41] in nine sequential steps (Figure 4). In the first step, all rock sample entries published before 1985 [12] were deleted because analytical methods—the way the chemical composition was analyzed—prior to this year are considered unreliable. This resulted in a new data matrix with the dimensions of  $5506 \times 181$ . In the second step, 39 of the first 41 features were deleted. The 'LOCATION' feature was preserved, as it was essential for creating the classes in the eighth step. The 'CITATION' feature was used to select the validation data in the ninth step. In this way, the data matrix was reduced to  $5506 \times 142$ . The first feature represents the location and the second represents the scientific paper where the sample's chemical composition was published, while the other 140 features contain the geochemical data of the rock samples. The chemical composition is expressed as the mass fraction (wt%) of the major elements, trace elements in parts per million (ppm) or parts per billion (ppb), and isotopes as ratios. In the third step, locations outside the research area were deleted, resulting in a  $4153 \times 142$  data matrix.



**Figure 4.** Block diagram of the data preparation steps.

Moreover, in the initial dataset, many NAN fields were present, which could be computed by the MLAs; therefore, in the fourth step, these were replaced with zeroes and, consequently, no data matrix dimension change occurred.

The chemical compositions provided by GEOROC were determined using the XRF method. Depending on the methodology used, the iron oxides' mass fraction (wt%) was provided either in FeO, FeOT, Fe<sub>2</sub>O<sub>3</sub>, or Fe<sub>2</sub>O<sub>3</sub>T. Thus, in the fifth step, to calculate the mass fraction uniformly, these four features were unified into one (Fe<sub>2</sub>O<sub>3</sub>T) through the following procedure: all oxides measured as FeOT were transformed to Fe<sub>2</sub>O<sub>3</sub>T by first multiplying the wt% of FeOT's chemical composition by 1.1111. The chemical composition of FeO was multiplied by 1.1111 to obtain its chemical composition in Fe<sub>2</sub>O<sub>3</sub>, then, this

value was added to the initial value of  $\text{Fe}_2\text{O}_3$ , resulting in the  $\text{Fe}_2\text{O}_3\text{T}$ . In this way, the table's new dimensions were  $4153 \times 139$ .

In the sixth step, features where zeroes represented more than 65.29% (this value corresponds to the oxide  $\text{P}_2\text{O}_5$ ) were deleted. This was necessary, as features beyond this threshold had zero values higher than 90%, making them useless for the training process of a ML model. This produced a  $4153 \times 12$  data matrix. In the seventh stage, 199 entries had to be deleted as all 10 of their geochemical data fields were equal to zero. The emerging new data matrix's dimensions were  $3954 \times 12$ . In the eighth step, the classes were prepared based on the four volcanic groups of the HVA [1,2]. Methana along with Aigina and Paros were labeled as Class 1. Melos together with Antimelos and Kimolos were labeled as Class 2. Santorini with Christianoi and Kolumbo were labeled as Class 3, and finally, Nisyros with Giali, and Pergoussa were labeled as Class 4.

In the final ninth step, 40 entries were randomly handpicked from the dataset to compile the validation dataset. This was done with the following constrains: 10 samples from each class, with each sample from a different publication and year. Furthermore, the 'CITATION' feature was removed, as it was no longer needed. This resulted in a  $3914 \times 11$  matrix dimension for the final dataset.

Following data reduction, six datasets (All datasets are available upon request) were created. The initial dataset was labeled Dataset 1 and consisted of 3914 entries and 11 features. The first 10 features were the geochemical data of the major elements and the last was the class. A snapshot of Dataset 1, the initial dataset, is shown in Table 1. These 3914 data samples were used to plot  $\text{SiO}_2$  vs.  $\text{K}_2\text{O}$  [42] (Figure 5) and  $\text{SiO}_2$  vs.  $\text{Na}_2\text{O} + \text{K}_2\text{O}$  [43] (Figure 6) to check if the four volcanic regions could be discriminated by the use of this bigger dataset. The diagrams show that this was not the case, as major overlapping of the sample points was present. In Dataset 2, the zeroes were replaced by the mean value. In Dataset 3, the zeroes were replaced by the median. Datasets 4–6 were created from Datasets 1–3 correspondingly by deleting the outliers. The outliers were removed by using the stats class of the Scipy library and the Numpy library. The number of entries of each class in every dataset is shown in Table 2. The basic statistic data of Datasets 1–6 are presented in Tables 3–8. The histograms of Datasets 1–3 with outliers are depicted in Figure 7. The histograms of Datasets 4–6 with the outliers removed are depicted in Figure 8. The histogram of Dataset 5 shows the best distribution of the data.

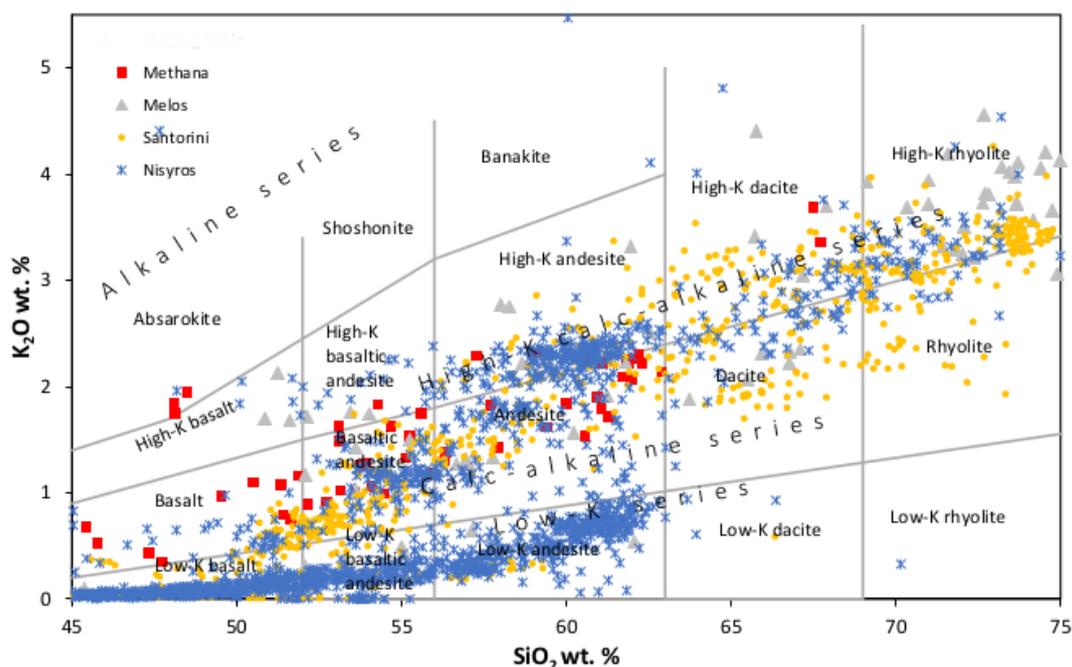


Figure 5.  $\text{K}_2\text{O}$  vs.  $\text{SiO}_2$  diagram.

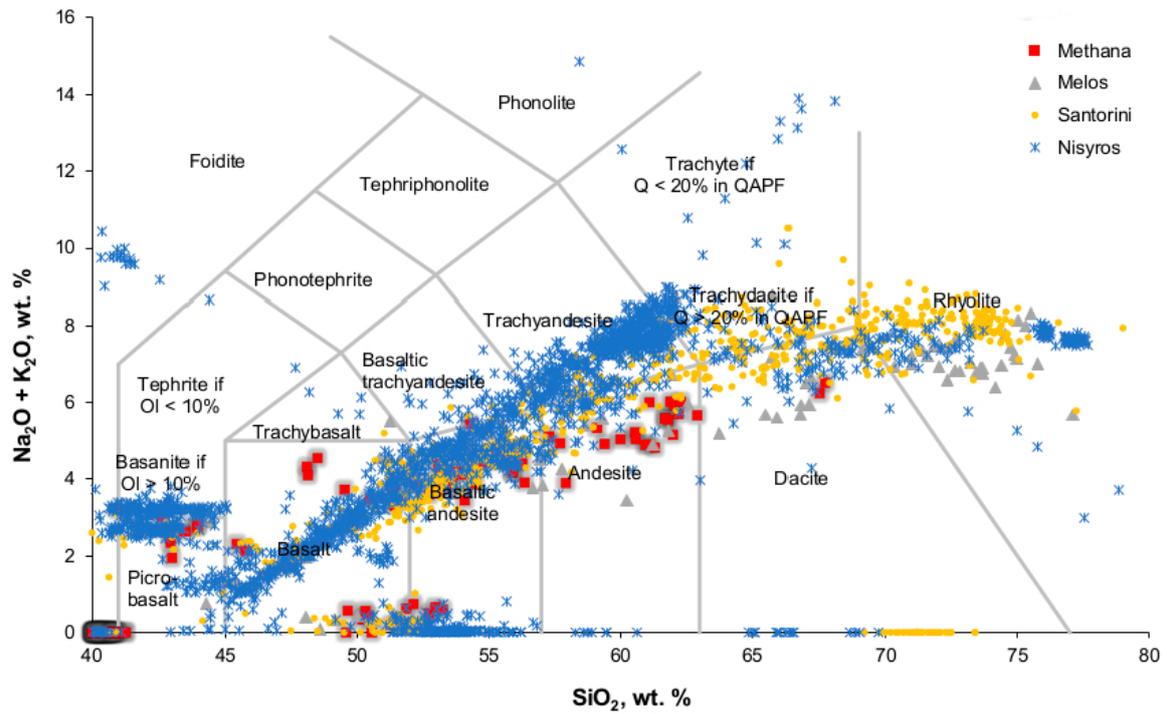


Figure 6. Na<sub>2</sub>O + K<sub>2</sub>O vs. SiO<sub>2</sub> diagram.

Table 1. Snapshot of Dataset 1.

Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10	Class
SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>	
67.72	0.32	15.7	2.75	3.63	1.23	0.07	3.36	3.16	0.11	1
67.5	0.36	15.15	3.30	4.85	1.69	0.08	3.69	2.55	0.17	1
60.55	0.52	17.26	5.61	6.26	2.49	0.11	2.25	2.98	0.14	1
71.58	0.25	12.72	2.07	1.49	0.5	0.06	4.19	3.01	0.03	2
69.13	0.38	14.06	2.79	2.23	0.67	0.09	3.93	3.47	0.07	2
71.41	0.38	14.05	2.56	2.73	0.77	0.04	3.32	4.02	0.08	2
67.89	0.39	14.11	2.84	2.43	0.64	0.11	3.70	3.73	0.08	2
54.28	0.84	16.97	8.11	8.52	5.56	0.15	1.57	3.07	0.12	3
52.21	0.84	17.79	8.70	10.81	6.17	0.16	0.60	2.71	0.10	3
52.3	0.85	17.67	8.52	10.73	6.13	0.16	0.60	2.78	0.11	3
55.31	0.92	19.19	7.18	8.55	4.11	0.12	1.09	3.61	0.20	4
55.52	0.92	18.99	7.15	8.51	4.30	0.12	1.09	3.42	0.22	4
54.91	0.92	19.51	7.17	8.61	4.34	0.12	1.08	3.61	0.19	4

Table 2. Classes and entries of Datasets 1–6.

	Methana (Class 1)	Melos (Class 2)	Santorini (Class 3)	Nysiros (Class 4)	Total
Dataset 1	274	115	1022	2503	3914
Dataset 2	274	115	1022	2503	3914
Dataset 3	274	115	1022	2503	3914
Dataset 4	135	104	908	2275	3422
Dataset 5	132	104	906	2263	3405
Dataset 6	130	104	905	2263	3402

**Table 3.** Basic statistics of Dataset 1.

	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	50.62	1.77	15.18	11.33	7.01	8.18	0.18	1.05	2.96	0.23
median	53.43	0.33	15.52	6.44	6.09	1.33	0.12	0.43	3.12	0
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max	81.58	50.47	68.35	100.42	53.85	52.47	1.81	9.62	9.92	44.94
std	17.47	6.34	10.79	18.41	6.48	13.44	0.26	1.40	2.25	2.70
var	305.37	40.24	116.33	339.1	42.04	180.73	0.07	1.96	5.08	7.27

**Table 4.** Basic statistics of Dataset 2.

	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	52.33	2.39	16.71	11.57	7.38	9.52	0.23	1.30	3.49	0.38
median	53.43	1.35	15.52	6.80	7.01	4.53	0.18	1.05	3.12	0.23
min	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
max	81.58	50.47	68.35	100.42	53.85	52.47	1.81	9.62	9.92	44.94
std	14.70	6.22	9.53	18.34	6.27	12.96	0.24	1.28	1.79	2.69
var	216.00	38.75	90.8	336.39	39.32	168.01	0.06	1.63	3.22	7.21

**Table 5.** Basic statistics of Dataset 3.

	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	52.43	1.89	16.74	11.46	7.33	8.40	0.22	1.15	3.52	0.34
median	53.43	0.33	15.52	6.44	6.10	1.33	0.12	0.43	3.12	0.17
min	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
max	81.58	50.47	68.35	100.42	53.85	52.47	1.81	9.62	9.92	44.94
std	14.69	6.31	9.52	18.36	6.28	13.32	0.24	1.33	1.79	2.69
var	215.94	39.85	90.70	336.92	39.39	177.41	0.06	1.78	3.19	7.22

**Table 6.** Basic statistics of Dataset 4.

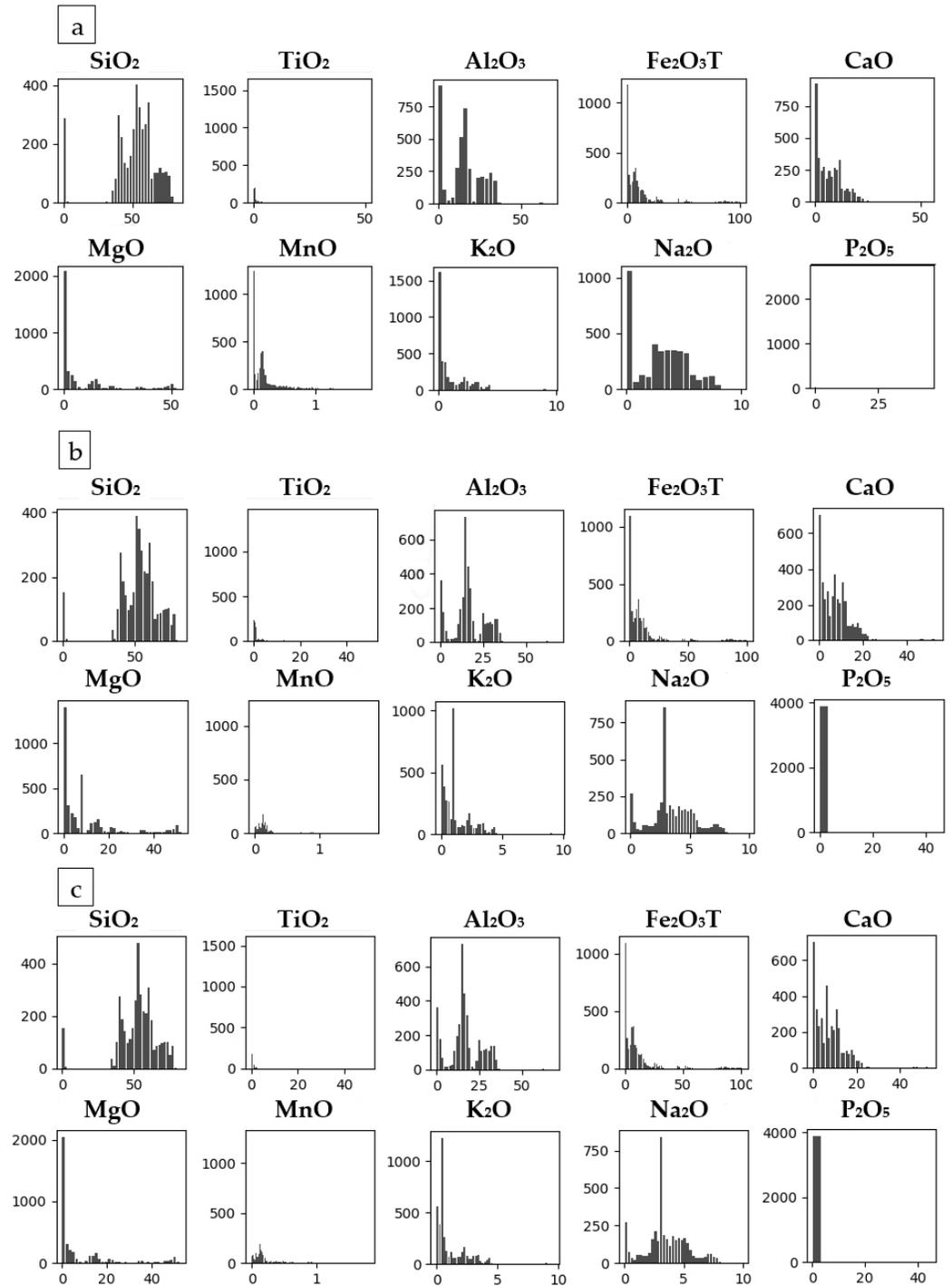
	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	54.58	0.6	16.93	6.83	7.75	6.6	0.13	1.11	3.36	0.06
median	54.72	0.34	16.5	4.98	7.21	1.20	0.10	0.55	3.45	0.00
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max	81.58	7.04	36.91	46.88	25.44	48.48	0.96	5.18	8.25	2.68
std	12.16	0.84	9.87	7.94	5.75	11.24	0.17	1.24	2.1	0.11
var	147.98	0.7	97.4	63.08	33.04	126.31	0.03	1.53	4.43	0.01

**Table 7.** Basic statistics of Dataset 5.

	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	55.44	1.23	18.00	6.97	7.95	7.98	0.19	1.28	3.67	0.20
median	54.79	1.09	16.53	5.53	7.27	4.33	0.18	1.05	3.46	0.23
min	30.42	0.00	0.00	0.02	0.08	0.00	0.00	0.00	0.00	0.00
max	81.58	7.04	36.91	46.88	25.44	48.4	0.94	4.81	8.25	2.68
std	10.02	0.82	8.77	7.62	5.61	10.73	0.14	1.14	1.79	0.07
var	100.41	0.66	76.92	58.02	31.48	115.05	0.02	1.31	3.19	0.01

**Table 8.** Basic statistics of Dataset 6.

	Feat.#1	Feat.#2	Feat.#3	Feat.#4	Feat.#5	Feat.#6	Feat.#7	Feat.#8	Feat.#9	Feat.#10
	SiO <sub>2</sub>	TiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub> T	CaO	MgO	MnO	K <sub>2</sub> O	Na <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>
mean	55.49	0.72	18.03	6.84	7.94	6.71	0.17	1.18	3.68	0.06
median	54.8	0.34	16.54	5.52	7.28	1.33	0.12	0.56	3.47	0.00
min	30.42	0.00	0.00	0.02	0.08	0.00	0.00	0.00	0.00	0.00
max	81.58	7.04	36.91	46.88	25.44	48.35	0.94	4.81	8.25	2.68
std	10.00	0.77	8.76	7.58	5.61	10.98	0.15	1.18	1.78	0.11
var	100.05	0.59	76.78	57.51	31.52	120.51	0.02	1.40	3.17	0.01



**Figure 7.** Histograms of the datasets with outliers: (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3.

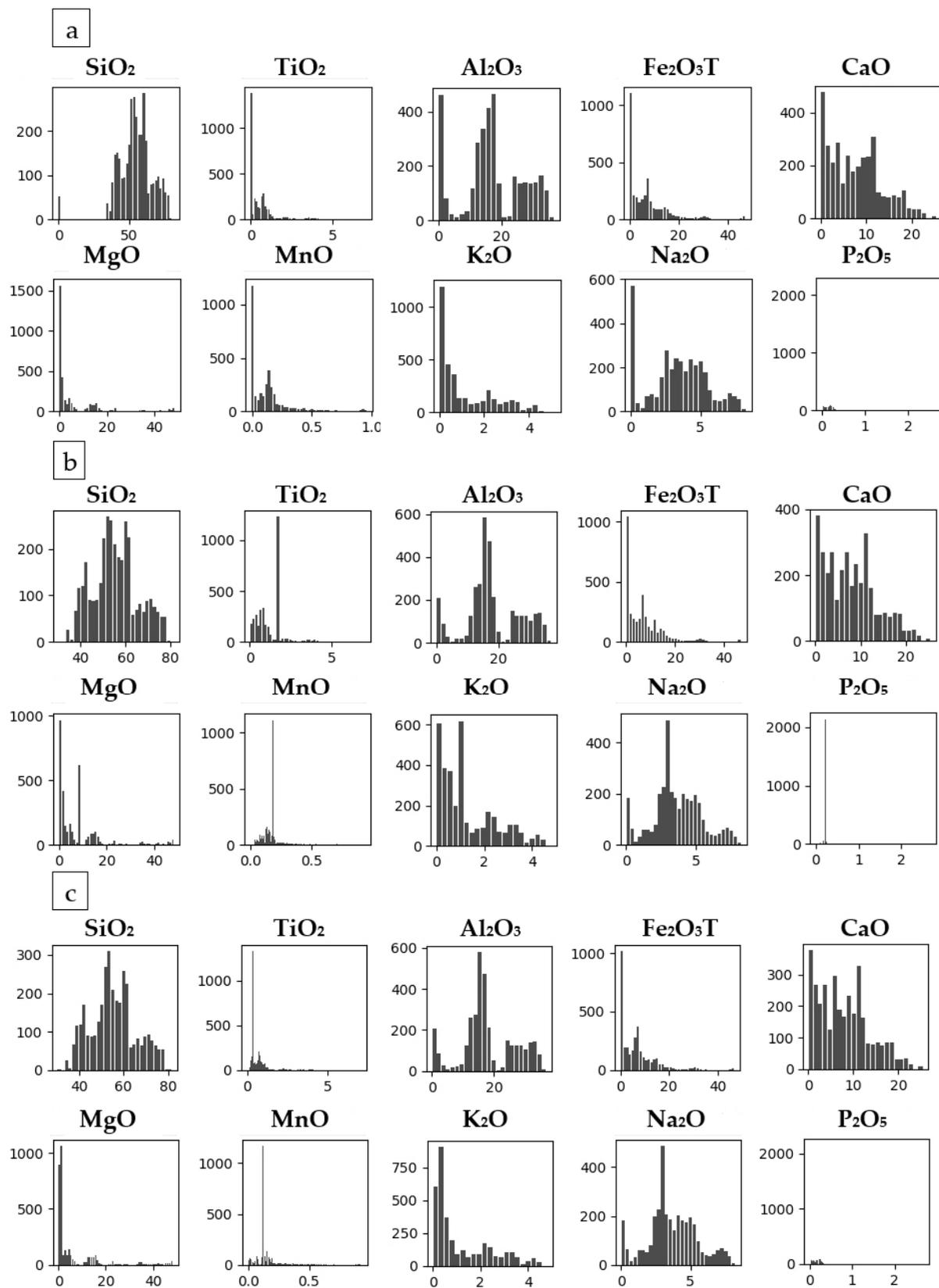


Figure 8. Histograms of the datasets after outlier removal: (a) Dataset 4, (b) Dataset 5, and (c) Dataset 6.

### 2.3. Simulations

The training of the models was implemented using Python 3.8 (Guido van Rossum, The Hague, The Netherlands) [44] joint with Scikit-learn (David Cournapeau, Paris, France) [45], the xGboost wrapper for Scikit-learn(<https://xgboost.readthedocs.io/>, accessed on 30 June 2021) and elm.ELM Kernel (Augusto Almeida, Brazil). The goal was to create the most accurate model, with the smallest standard deviation and the least computing time, by using one out of the 10 MLAs and one of the six datasets. Training of the model was conducted on a Windows 10 System with an Intel Core 2 Quad Q8400 CPU and 8 GB of RAM. Tenfold cross-validation was applied (i.e., the original sample was randomly partitioned into 10 equal size subsamples). The datasets were split randomly into 70% training data and 30% testing data. In this way, the experiments were repeated 10 times by using 10 different versions of the dataset. Finally, three stages of testing and three improvement procedures were applied (Figure 9).

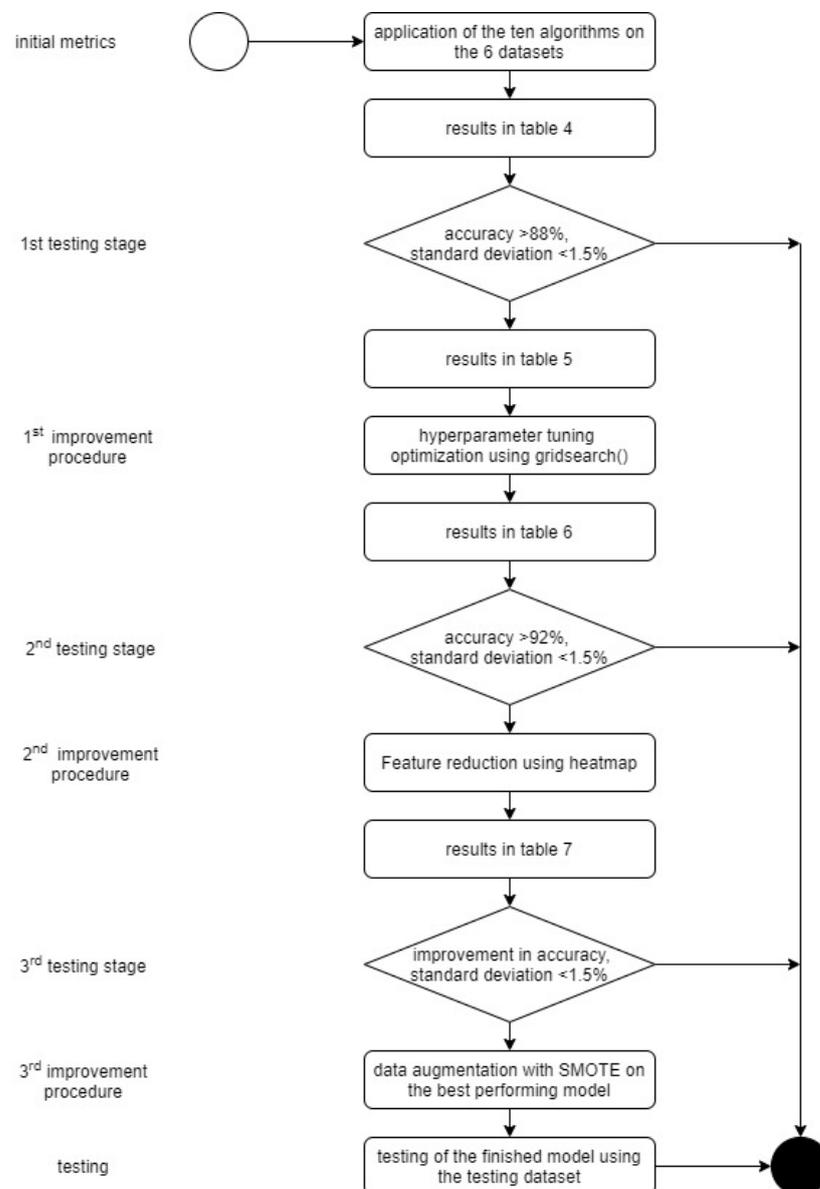


Figure 9. Flowchart of the model’s development.

The initial metrics were computed by using the six datasets on the 10 algorithms with their default parameters.

The threshold values for the first testing stage were an accuracy of >88% and a standard deviation of <1.5%. In the following initial improvement procedure, a grid search was applied to the MLAs of the models passing the first testing stage to optimize the selected model's algorithms.

The second testing stage had a threshold accuracy of >92% and the standard deviation remained at <1.5%. The models outperforming all the others were then put into the second improvement procedure, where feature reduction was applied using the results of a heatmap.

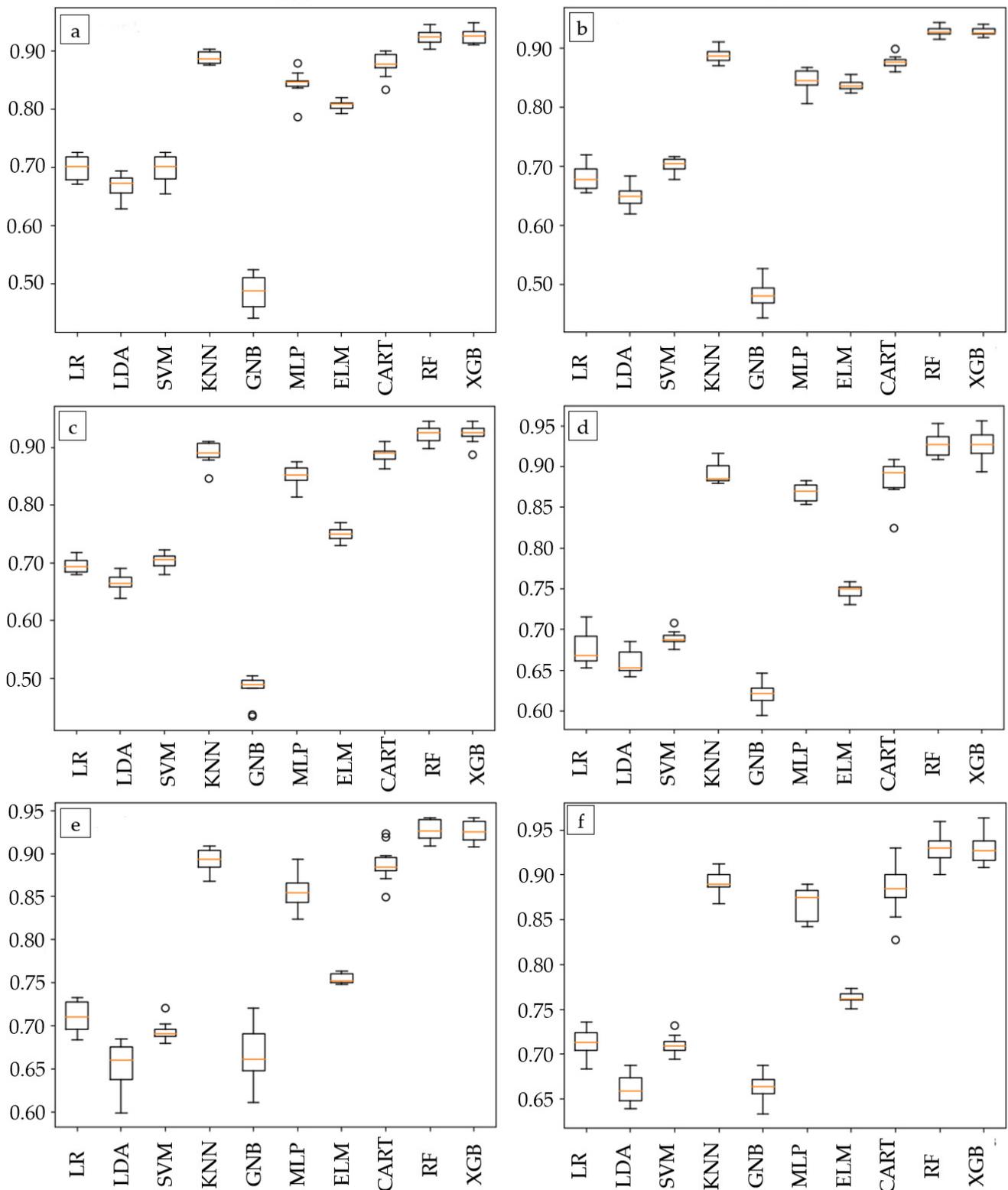
In the third testing stage, the goal was to improve accuracy by keeping the standard deviation of <1.50%. Following this, a confusion matrix [46] was compiled for the best-performing model. This revealed that a third improvement procedure was necessary to augment the data. Namely, the Synthetic Minority Over-sampling Technique (SMOTE) [47] was applied to the best-performing model's dataset. Finally, the best-performing model was tested using the testing data, i.e., the 40 entries left out of the training and validation.

### 3. Results

The results of the first testing stage showed accuracy ranging from 48.39% (2.24% standard deviation, 2.01 s) for the model using GNB with Dataset 2 to 92.75% (1.36% standard deviation, 11.48 s) for the model using XGB with Dataset 5 (Table 9, Figure 10).

**Table 9.** Testing results for the first testing stage (LR: Logistic Regression, LDA: Linear Discriminant Analysis, SVM: Support Vector Machine, KNN: K Nearest Neighbor, GNB: Gaussian Naïve Bayes, MLP: Multi-Layer Perceptron, ELM: Extreme Learning Machine, CART: Classification and Regression Trees, RF: Random Forest, XGB: xGBoost, std: standard deviation).

Dataset	LR			LDA			SVM			KNN		
	Accuracy	Std	Time (Sec)									
1	69.82%	2.15%	2.66	66.84%	2.02%	2.34	69.75%	2.23%	5.05	88.02%	1.97%	3.49
2	68.03%	1.85%	2.7	64.74%	2.18%	2.2	70.20%	1.64%	5.32	87.77%	1.82%	3.66
3	68.64%	1.87%	2.68	65.98%	2.32%	2.14	70.17%	1.75%	5.43	88.57%	1.38%	4.14
4	67.67%	2.08%	2.52	65.99	1.47%	2.32	68.94%	0.85%	3.88	88.43%	1.91%	3.6
5	71.00%	1.74%	2.52	65.42%	2.54%	2.32	69.31%	1.12%	5.25	87.96%	1.57%	3.86
6	71.22%	1.60%	2.48	66.12%	1.60%	2.45	71.00%	1.06%	5.42	88.58%	1.08%	3.77
Dataset	GNB			MLP			ELM			CART		
	Accuracy	Std	Time (Sec)									
1	48.48%	2.89%	2.00	84.64%	1.80%	30.00	80.46%	1.30%	3.41	87.37%	2.18%	3.63
2	48.39%	2.24%	2.01	87.61%	1.74%	30.12	85.52%	1.50%	0.95	87.66%	1.48%	3.59
3	48.74%	2.45%	2.17	83.65%	1.86%	30.42	74.27%	3.45%	1.44	87.66%	1.62%	3.61
4	62.07%	1.60%	2.01	86.33%	0.98%	22.59	74.78%	5.22%	0.44	88.02%	1.89%	3.41
5	66.30%	3.32%	2.03	85.72%	2.03%	31.22	77.54%	1.78%	1.35	87.58%	2.30%	3.6
6	66.19%	1.60%	2.21	86.81%	1.77%	32.35	76.13%	1.92%	1.01	87.86%	2.12%	3.59
Dataset	RF			XGB								
	Accuracy	Std	Time (Sec)	Accuracy	Std	Time (Sec)						
1	91.60%	1.60%	10.3	91.57%	1.34%	13.19						
2	91.68%	1.76%	10.18	91.42%	1.60%	13.02						
3	91.64%	1.24%	10.39	91.97%	1.55%	13.02						
4	92.02%	1.54%	9.64	91.82%	1.99%	11.44						
5	92.73%	1.46%	9.26	92.75%	1.36%	11.48						
6	92.69%	1.48%	9.05	92.73%	1.68%	11.33						



**Figure 10.** Boxplots of the simulation results: (a) Dataset 1, (b) Dataset 2, (c) Dataset 3, (d) Dataset 4, (e) Dataset 5, and (f) Dataset 6.

Six models met the requirements of the first testing stage (>88% accuracy and <1.5% standard deviation). These were Model 1: kNN combined with Dataset 6 (88.58% accuracy, 1.08% standard deviation, 3.77 s); Model 2: RF with Dataset 3 (91.64% accuracy, 1.24% standard deviation, 10.39 s); Model 3: RF with Dataset 5 (92.73% accuracy, 1.46% standard

deviation, 9.64 s); Model 4: RF with Dataset 6 (92.69% accuracy, 1.48% standard deviation, 9.05 s); model 5: XGB combined with Dataset 1 (91.57% accuracy, 1.34% standard deviation, 13.19 s); and Model 6: XGB combined with Dataset 5 (92.75% accuracy, 1.36% standard deviation, 11.48 s). All MLAs were used with their standard parameters (Table 10).

**Table 10.** Best performing models (first testing stage).

Model	MLA	Dataset	Accuracy	Std	Time (Sec)	Parameters
1	kNN	6	88.58%	1.08	3.52	n_neighbors = 5, weight = uniform, metric = Euclidean
2	RF	3	91.64%	1.24%	10.39	max depth = 3, n_estimators = 100
3	RF	5	92.73%	1.46%	9.64	max depth = 3, n_estimators = 100
4	RF	6	92.69%	1.48%	9.05	max depth = 3, n_estimators = 100
5	XGB	1	91.57%	1.34%	13.19	max depth = 3, n_estimators = 100, learning_rate = 0.1
6	XGB	5	92.75%	1.36%	11.48	max depth = 3, n_estimators = 100, learning_rate = 0.1

The performance of a MLA can be optimized by changing a set of parameters. Each MLA has its own parameter that can be fine-tuned. The six models were improved in the first improvement procedure by using the gridsearch() function [45], which allows bulk testing of models on a selected set of parameters. This procedure is known as hyperparameter tuning. In the second testing stage (accuracy >92%, standard deviation <1.5%), only three models reached these stricter criteria, namely Model 3: RF with Dataset 5 (92.79% accuracy, 1.29% standard deviation, 16.13 s); Model 4: RF with Dataset 6 (92.75% accuracy, 1.35% standard deviation, 15.91 s); and Model 5: XGB combined with Dataset 5 (92.80% accuracy, 1.48% standard deviation, 18.97 s). A side effect of the improvement in accuracy was a rise in computational time during the training (Table 11).

**Table 11.** Performance improvement of the models after hyperparameter tuning (first improvement procedure).

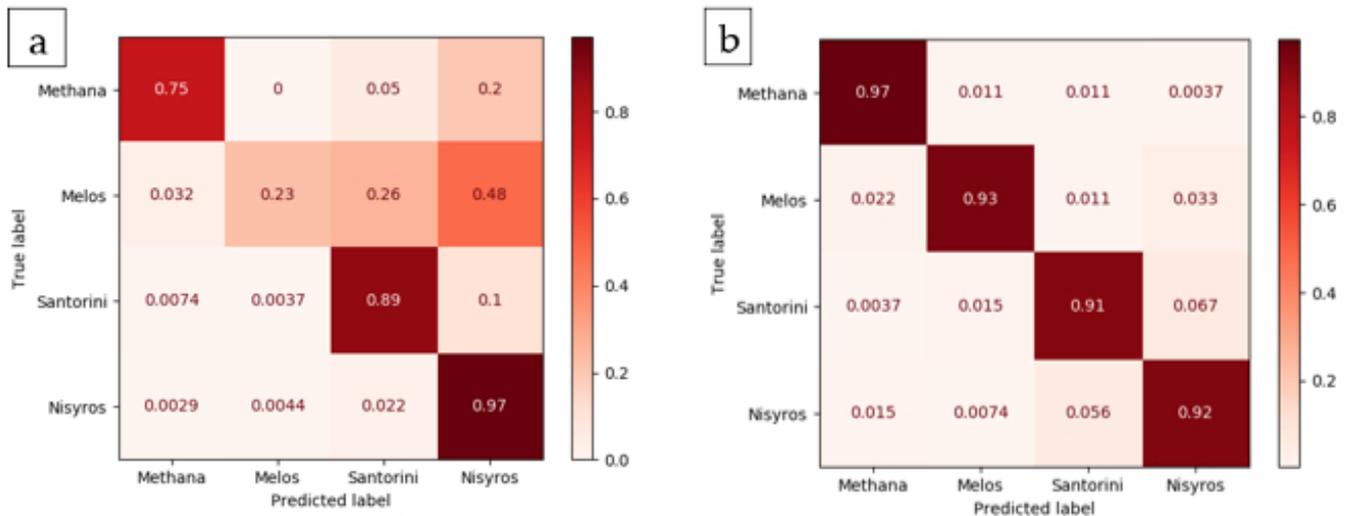
Model	MLA	Dataset	Accuracy	Std	Time (Sec)	Parameters
1	kNN	6	91.26%	1.70%	3.21	n_neighbors = 2, weights = distance, metric = Manhattan
2	RF	3	91.85%	0.95%	30.16	max depth = 26, n_estimators = 200
3	RF	5	92.79%	1.29%	16.13	max depth = 17, n_estimators = 200
4	RF	6	92.75%	1.35%	15.91	max depth = 14, n_estimators = 200
5	XGB	1	91.93%	1.56%	23.03	max depth = 9, n_estimators = 106, learning_rate = 0.1
6	XGB	5	92.80%	1.48%	18.97	max depth = 8, n_estimators = 101, learning_rate = 0.1

In the second improvement procedure, the best models from the second testing stage were analyzed to accomplish feature reduction. A heatmap was used to rank the features of Datasets 5 and 6 and further improve Models 3, 4, and 6. The features of Dataset 5 were reordered as follows: Al<sub>2</sub>O<sub>3</sub>, TiO<sub>2</sub>, CaO, P<sub>2</sub>O<sub>5</sub>, Na<sub>2</sub>O, MnO, SiO<sub>2</sub>, Fe<sub>2</sub>O<sub>3</sub>T, MgO, and K<sub>2</sub>O. The features of Dataset 6 were reordered as follows: Al<sub>2</sub>O<sub>3</sub>, CaO, Na<sub>2</sub>O, TiO<sub>2</sub>, MnO, P<sub>2</sub>O<sub>5</sub>, SiO<sub>2</sub>, Fe<sub>2</sub>O<sub>3</sub>T, K<sub>2</sub>O, and MgO. After using all the rearranged features, the three models showed minor improvements (Table 12). Model 6 achieved the best results. Rearranging the features resulted in an improvement in computational time. By removing features based on the aforementioned ranking, no increase in accuracy was achieved.

**Table 12.** Performance improvement of the models after applying a heatmap (second testing stage, second improvement procedure).

Model	MLA	Dataset	Accuracy	Std	Time (Sec)	Parameters
3	RF	6	92.81%	1.3%	15.28	max depth = 17, n_estimators= 200
4	RF	5	92.90%	1.25%	14.43	max depth = 14, n_estimators= 200
6	XGB	5	92.94%	1.15%	13.85	max depth = 8, n_estimators= 101, learning_rate = 0.1

In the third improvement procedure, a confusion matrix was created for Model 6 to evaluate the accuracy of each class (Figure 11a). The low prediction accuracy results of Melos can be explained if we take into account that an extremely imbalanced dataset was used: 132 entries for Methana, 104 for Melos, 906 for Santorini, and 2263 for Nisyros (Table 2).



**Figure 11.** Confusion matrix of (a) Model 6 and (b) Model 6 with the augmented Dataset 5—balanced 900.

For this reason, SMOTE was applied to augment Dataset5. One new dataset was created, labeled Dataset 5—balanced 900. The confusion matrix (Figure 11b) showed an improvement. The final model was devised by using the optimized XGB algorithm and Dataset 5—balanced 900 created by this imputation. Accuracy improved to 93.07% and the standard deviation was calculated as 1.21%.

This final model was tested with the 40 handpicked entries left out of the training and testing data. In this validation sample, 90% accuracy was achieved. The Methana class scored 100% and the Melos class scored 80%. One sample was misclassified as Santorini and one as Nisyros. The Santorini class scored 90%, with only one sample being misclassified as Melos. The Nisyros class scored 90%, one sample being misclassified as Santorini.

#### 4. Discussion

The  $\text{SiO}_2$  vs.  $\text{K}_2\text{O}$  diagram plotted in this study using 3914 data points showed that it is not possible to discriminate the tectonic regions by using this methodology. This is in accordance with the  $\text{SiO}_2$  vs.  $\text{K}_2\text{O}$  diagrams of the HVA's volcanic regions using 1500 data points in 2005 [1]. In order to classify the volcanic regions, the simulations showed that the best result was achieved by Model 6, which used the XGB classifier and a dataset where imputation was applied by replacing the zero values with the mean value and where the outliers were removed (Dataset 5). This final model performed at an accuracy of 92.75%. After hyperparameter tuning, feature rearrangement, and data augmentation had been applied, the accuracy reached 93.07%. This shows that this procedure can augment accuracy and should always be applied. It is notable to mention that Model 1, which used the kNN classifier with Dataset 6, performed at an accuracy of 91.26% with a very low computational cost of only 3.21 s. The final validation was conducted using the validation dataset consisting of 40 entries left out of the training and testing which resulted in an accuracy of 90%.

This study is in accordance with the positive results obtained from the Calabrian Volcanic Arc [21]. In this study, a similar dataset was used to discriminate the volcanic groups of Central Italy in the framework of tephrochronological studies using the SVM classifier. In this research, 98% accuracy was achieved.

The use of the model developed in the present study should not be used as a black box. Other means of examining the rock samples, such as mineralogy and the age of crystallization, must be taken into account to cross-validate the model's results. The misclassification of four samples of the validation dataset shows that further improvement of the methodology described is necessary. This will be possible by retraining the selected classifier with new data which will emerge in the future.

## 5. Conclusions

The results of this research prove that the ML methodology used for discriminating tectonic environments can also be used to discriminate volcanic regions belonging to the same tectonic setting. In the case of the present study, four volcanic regions of the HVA producing IABs in a converging tectonic environment were successfully discriminated using the XGB classifier with an accuracy of 93.07% during training and 90% during validation.

The final model compiled was used for the implementation of a CBA [48] written in Python code and deployed in the flask framework. The CBA implemented in the framework of this study can be used as a valuable tool in the field of geology to classify volcanic rocks found in the Southern Aegean Sea in one of the four active volcanic regions of the HVA using 10 major elements. This will also be a useful tool for mapping of the HVA in the future. Furthermore, in geoarchaeology, the CBA can be used as a valuable tool for tracking the origin of artifacts found all over Greece, such as millstones and anchors made from volcanic rock dating from prehistory to antiquity, a procedure known as geochemical fingerprinting. The CBA is publicly accessible at [www.geology.gr](http://www.geology.gr) (accessed on 6 September 2021).

The authors' future plans include applying this methodology for the two active volcanic arcs in the Mediterranean, namely the Hellenic Volcanic Arc in the Aegean Sea and the Calabrian Volcanic Arc in the Tyrrhenian Sea, as similar datasets are available in the GEOROC database. This would produce a tool for applying geochemical fingerprinting of ancient artifacts to a broader geographic region.

**Author Contributions:** Conceptualization, A.G.O. and G.A.P.; methodology, A.G.O. and G.A.P.; software, A.G.O.; validation, A.G.O. and G.A.P.; investigation, A.G.O.; writing—original draft preparation, A.G.O.; writing—review and editing, A.G.O. and G.A.P.; visualization, A.G.O.; supervision, G.A.P.; project administration, G.A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The Geochemical Rock Database used in this study was retrieved from the Max Planck Institute for Chemistry, <http://georoc.mpch-mainz.gwdg.de/georoc/> accessed on 30 June 2021.

**Acknowledgments:** We would like to pay our respects to all the individuals involved in collecting rock samples in the field, those in the laboratory producing the geochemical data, and, last but not least, the people responsible for GEOROC preparing the datasets and distributing them for free. This research would not have been possible without them. Moreover, we wish to express our gratitude to the geologists Katrivanos Manolis and Ntarlagiannis Nikos for reviewing the geological context of this paper. Finally, we would like to thank Michailidou Vasiliki for linguistic revision of the manuscript.

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

## References

1. Francalanci, L.; Vougioukalakis, G.E.; Perini, G.; Manetti, P. A West-East Traverse along the Magmatism of the South Aegean Volcanic Arc in the Light of Volcanological, Chemical and Isotope Data. *Dev. Volcanol. Elsevier* **2005**, *7*, 65–111. [[CrossRef](#)]
2. Fytikas, M.; Innocenti, F.; Manetti, P.; Peccerillo, A.; Mazzuoli, R.; Villari, L. Tertiary to Quaternary Evolution of Volcanism in the Aegean Region. *Geol. Soc. Lond. Spec. Publ.* **1984**, *17*, 687–699. [[CrossRef](#)]
3. Pe-Piper, G.; Piper, D.J.W. The South Aegean Active Volcanic Arc: Relationships between Magmatism and Tectonics. *Dev. Volcanol. Elsevier* **2005**, *7*, 113–133. [[CrossRef](#)]
4. Peccerillo, A.; Taylor, S.R. Geochemistry of eocene calc-alkaline volcanic rocks from the Kastamonu area, Northern Turkey. *Contr. Mineral. Petrol.* **1976**, *58*, 63–81. [[CrossRef](#)]
5. Pearce, J.A.; Cann, J.R. Tectonic Setting of Basic Volcanic Rocks Determined Using Trace Element Analyses. *Earth Planet Sci. Lett.* **1973**, *19*, 290–300. [[CrossRef](#)]
6. Bhatia, M.R. Plate Tectonics and Geochemical Composition of Sandstone. *J. Geol.* **1983**, *91*, 611–627. [[CrossRef](#)]
7. Snow, C.A. A Reevaluation of Tectonic Discrimination Diagrams and a New Probabilistic Approach Using Large Geochemical Databases: Moving beyond Binary and Ternary Plots. *J. Geophys. Res.* **1973**, *111*, B06206. [[CrossRef](#)]
8. Vermeesch, P. Tectonic Discrimination Diagrams Revisited. *Geochem. Geophys. Geosystems* **2006**, *7*. [[CrossRef](#)]
9. Zhang, Q.; Sun, W.; Zhao, Y.; Yuan, F.; Jiao, S.; Chen, W. New Discrimination Diagrams for Basalts Based on Big Data Research. *Big Earth Data* **2019**, *3*, 44–45. [[CrossRef](#)]
10. Max Planck Institute for Chemistry. Geochemical Rock Database. Available online: <http://georoc.mpch-mainz.gwdg.de/georoc/> (accessed on 29 June 2021).
11. Lehnert, K.; Robinson, S. EarthChem. Available online: <https://www.earthchem.org/petdb> (accessed on 29 June 2021).
12. Li, C.; Arndt, N.T.; Tang, Q.; Ripley, E.M. Trace Element Indiscrimination Diagrams. *Lithos* **2015**, *232*, 76–83. [[CrossRef](#)]
13. Han, S.; Li, M.; Ren, Q.; Chengzhao, L. Intelligent Determination and Data Mining for Tectonic Settings of Basalts Based on Big Data Methods. *Acta Petrol. Sin.* **2018**, *34*, 3207–3216.
14. Papakostas, G.A.; Diamantaras, K.I.; Palmieri, F.A. Emerging Trends in Machine Learning for Signal Processing. *Comput. Intell. Neurosci.* **2017**, *2017*, 1–2. [[CrossRef](#)] [[PubMed](#)]
15. Vermeesch, P. Tectonic Discrimination of Basalts with Classification Trees. *Geochim. Et Cosmochim. Acta* **2006**, *70*, 1839–1848. [[CrossRef](#)]
16. Petrelli, M.; Perugini, D. Solving Petrological Problems through Machine Learning: The Study Case of Tectonic Discrimination Using Geochemical and Isotopic Data. *Contrib. Miner. Pet.* **2016**, *171*, 81. [[CrossRef](#)]
17. Ueki, K.; Hino, H.; Kuwatani, T. Geochemical Discrimination and Characteristics of Magmatic Tectonic Settings: A Machine-Learning-Based Approach. *Geochem. Geophys. Geosyst.* **2018**, *19*, 1327–1347. [[CrossRef](#)]
18. Ren, Q.; Li, M.; Han, S. Tectonic Discrimination of Olivine in Basalt Using Data Mining Techniques Based on Major Elements: A Comparative Study from Multiple Perspectives. *Big Earth Data* **2019**, *3*, 8–25. [[CrossRef](#)]
19. Ren, Q.; Li, M.; Han, S.; Zhang, Y.; Zhang, Q.; Shi, J. Basalt Tectonic Discrimination Using Combined Machine Learning Approach. *Minerals* **2019**, *9*, 376. [[CrossRef](#)]
20. Itano, K.; Ueki, K.; Iizuka, T.; Kuwatani, T. Geochemical Discrimination of Monazite Source Rock Based on Machine Learning Techniques and Multinomial Logistic Regression Analysis. *Geosciences* **2020**, *10*, 63. [[CrossRef](#)]
21. Petrelli, M.; Bizzarri, R.; Morgavi, D.; Baldanza, A.; Perugini, D. Combining Machine Learning Techniques, Microanalyses and Large Geochemical Datasets for Tephrochronological Studies in Complex Volcanic Areas: New Age Constraints for the Pleistocene Magmatism of Central Italy. *Quat. Geochronol.* **2017**, *40*, 33–44. [[CrossRef](#)]
22. Jordan, M.J.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
23. Mitchell, T.M. *Machine Learning*, 1st ed.; McGraw-Hill Education: New York, NY, USA, 1997.
24. Peng, C.-Y.J.; Lee, K.L.; Ingersoll, G.M. An Introduction to Logistic Regression Analysis and Reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [[CrossRef](#)]
25. Li, T.; Zhu, S.; Ogihara, M. Using Discriminant Analysis for Multi-Class Classification: An Experimental Investigation. *Knowl. Inf. Syst.* **2006**, *10*, 453–472. [[CrossRef](#)]
26. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inform. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
27. John, G.H.; Langley, P. Estimating Continuous Distributions in Bayesian Classifiers. *arXiv* **2013**, arXiv:1302.4964.
28. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
29. Collobert, R.; Bengio, S. Links between Perceptrons, MLPs and SVMs. In Proceedings of the Twenty-First international conference on Machine learning-ICML '04, Banff, AB, Canada, 20 September 2004; ACM Press: Banff, AB, Canada, 2004; p. 23. [[CrossRef](#)]
30. Minsky, M.; Papert, S. *Perceptrons*; M.I.T. Press: Oxford, UK, 1969.
31. Ding, S.; Xu, X.; Nie, R. Extreme Learning Machine and Its Applications. *Neural Comput. Appl.* **2014**, *25*, 549–556. [[CrossRef](#)]
32. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
33. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification And Regression Trees*, 1st ed.; Routledge: Boca Raton, FL, USA, 2017. [[CrossRef](#)]
34. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]

35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
36. Tin Kam Ho. Random Decision Forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*; Institute of Electrical and Electronics Engineers Computer Society Press: Montreal, QC, Canada, 1995; Volume 1, pp. 278–282. [[CrossRef](#)]
37. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Francisco, CA, USA, 2016; pp. 785–794. [[CrossRef](#)]
38. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Statist.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
39. Yan, X.; Su, X.G. *Linear Regression Analysis: Theory and Computing*; World Scientific Publishing Company: Singapore, 2009. [[CrossRef](#)]
40. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, Austin, TX, USA, 9–15 July 2010.
41. van der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [[CrossRef](#)]
42. Ewart, A. The Mineralogy and Petrology of Tertiary-Recent Orogenic Volcanic Rocks: With a Special Reference to the Andesitic-Basaltic Compositional Range. In *Andesites: Orogenic Andesites and Related Rocks*; Thorpe, R.S., Ed.; Wiley: Chichester, UK, 1982; pp. 25–95.
43. Bas, M.J.L.; Maitre, R.W.L.; Streckeisen, A.; Zanettin, B.; IUGS Subcommittee on the Systematics of Igneous Rocks. A Chemical Classification of Volcanic Rocks Based on the Total Alkali-Silica Diagram. *J. Petrol.* **1986**, *27*, 745–750. [[CrossRef](#)]
44. Van Rossum, G.; Drake, F.L., Jr. *Python Reference Manual*; Centrum voor Wiskunde en Informatica: Amsterdam, The Netherlands, 1995.
45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-Learn: Machine Learning in Python. *arXiv* **2018**, arXiv:1201.0490.
46. Kohavi, R.; Provost, F. Glossary of Terms. *Mach. Learn.* **1998**, *30*, 271–274. [[CrossRef](#)]
47. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intel. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
48. Ouzounis, A. Geology in Greece. Available online: <http://geology.gr/> (accessed on 6 September 2021).