CrossMark

ORIGINAL PAPER

Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data

Maurizio Petrelli[1] [ORCID] · **Diego Perugini[1]**

Abstract Machine-learning methods are evaluated to study the intriguing and debated topic of discrimination among different tectonic environments using geochemical and isotopic data. Volcanic rocks characterized by a whole geochemical signature of major elements (SiO_2, TiO_2, Al_2O_3, Fe_2O_{3T}, CaO, MgO, Na_2O, K_2O), selected trace elements (Sr, Ba, Rb, Zr, Nb, La, Ce, Nd, Hf, Sm, Gd, Y, Yb, Lu, Ta, Th) and isotopes ($^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{208}Pb/^{204}Pb$, $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$) have been extracted from open-access and comprehensive petrological databases (i.e., PetDB and GEOROC). The obtained dataset has been analyzed using support vector machines, a set of supervised machine-learning methods, which are considered particularly powerful in classification problems. Results from the application of the machine-learning methods show that the combined use of major, trace elements and isotopes allows associating the geochemical composition of rocks to the relative tectonic setting with high classification scores (93 %, on average). The lowest scores are recorded from volcanic rocks deriving from back-arc basins (65 %). All the other tectonic settings display higher classification scores, with oceanic islands reaching values up to 99 %. Results of this study could have a significant impact in other petrological studies potentially opening new perspectives for petrologists and geochemists. Other examples of applications include the development of more robust geothermometers and geobarometers and the recognition of

volcanic sources for tephra layers in tephro-chronological studies.

Keywords Machine learning · Large petrological databases · Tectonic discrimination · Major and trace elements · Isotopes

Introduction

Machine learning (ML) entails the use of algorithms and techniques to detect patterns from large datasets and to exploit the uncovered patterns to predict future trends, classify, or perform other kind of strategic decisions (Murphy 2012). The field of ML has progressed dramatically over the past two decades, developing from a "numerical curiosity" to a practical technology with widespread scientific and commercial use (Jordan and Mitchell 2015). For example, ML is now successfully utilized in several fields like computer vision, speech recognition, natural language processing and robot control (Jordan and Mitchell 2015).

In principle, each complex problem characterized by a large enough number of input samples is well suited for ML applications (Jordan and Mitchell 2015). It is notable that the application of ML techniques has been quite extensively tested in the Earth Sciences (Huang et al. 2002; Petrelli et al. 2003; Masotti et al. 2006; Cannata et al. 2011; Zuo and Carranza 2011; Abedi et al. 2012; Goldstein and Coco 2014) but, surprisingly, their use is still virtually unexplored with regards to the solution of petrological problems. One intriguing and debated petrological application, potentially well suited for the investigation by ML, is the tectonic discrimination of magmas using geochemical data (Li et al. 2015).

✉ Maurizio Petrelli
maurizio.petrelli@unipg.it

1 Department of Physics and Geology, University of Perugia, Piazza Università, 06100 Perugia, Italy

Springer

Trace element discrimination diagrams were introduced in the 1970s as a method for identifying the tectonic setting of basalts and other volcanic rocks (Pearce and Cann 1973). At that time, classification diagrams utilized only a few elements plotted as binary or triangular diagrams (Pearce and Cann 1973; Pearce 1976; Pearce and Norry 1979; Wood 1980; Shervais 1982; Meschede 1986; Grimes et al. 2015). This approach is still widely used; to date, the work by Pearce and Cann (1973) received about 2082 citations (more than 400 only in the last 5 years; source: Scopus, August 2016) testifying for the popularity of this approach in the petrological community (Li et al. 2015).

In 2006, Snow (2006) demonstrated that the success of these diagrams is mainly hindered by their limited dimensionality due to visualization requirements. Snow (2006) proposed alternative probabilistic methods and reported single analysis classification success rates for volcanic rocks from island arcs, ocean islands and mid-ocean ridges environments of about 83, 75 and 76 %, respectively.

In addition, Vermeesch (2006a) firstly proposed the application of two-dimensional linear discriminant analysis (LDA, Vermeesch 2006a) and the application of classification trees (Vermeesch 2006b) to statistically determine the tectonic affinity of oceanic basalts. LDA has also been implemented by other authors (Agrawal et al. 2004, 2008; Verma et al. 2013) for the solution of similar problems.

More recently, Li et al. (2015) further highlighted the inaccuracy of binary and ternary diagrams as discriminating tools in assigning tectonic settings starting from geochemical analyses of igneous rocks. As an alternative, Li et al. (2015) suggested the use of primordial mantle normalized diagrams as qualitative discriminating tools for the different tectonic environments.

In this study, we introduce ML basics and its potential in petrological and mineralogical applications. A ML algorithm is then applied in the attempt to discriminate magmas from different tectonic settings using geochemical data (major elements, trace elements and isotopic data) as input parameters. Although this is a long-standing, highly debated and controversial geological problem, a renewed interest in its solution has developed in recent times among petrologists (see e.g., Li et al. (2015) and references therein). The aim of our work is to cut the Gordian knot representing the complexity associated with the identification of tectonic environments starting from the geochemical composition of magmas.

To provide a robust and quantitative tool, we extract a large number of samples from open-access and comprehensive global petrological databases like PetDB (http://www.earthchem.org/petdb) and GEOROC (http://georoc.mpch-mainz.gwdg.de/georoc/). We also test different approaches involving major elements, trace elements and isotopes with the aim of defining the best strategy for the discrimination

among different tectonic settings. Finally, we highlight the drawbacks and limitations of the method, and we discuss our results in light of the published literature.

Methods

Basics of machine learning and potential applications to mineralogy and petrology

ML is one of the fastest growing areas of computer science, with far-reaching applications (Jordan and Mitchell 2015). One common feature for these applications is that, due to the complexity of the problems that need to be resolved, a human programmer cannot provide an explicit, fine-detailed univocal solution (Shai and Shai 2014). As a consequence, ML algorithms do not try to solve complex problems using an a priori defined conceptual model but they attempt to unravel complexities in large datasets through a so-called learning process (Bishop 2007; Shai and Shai 2014).
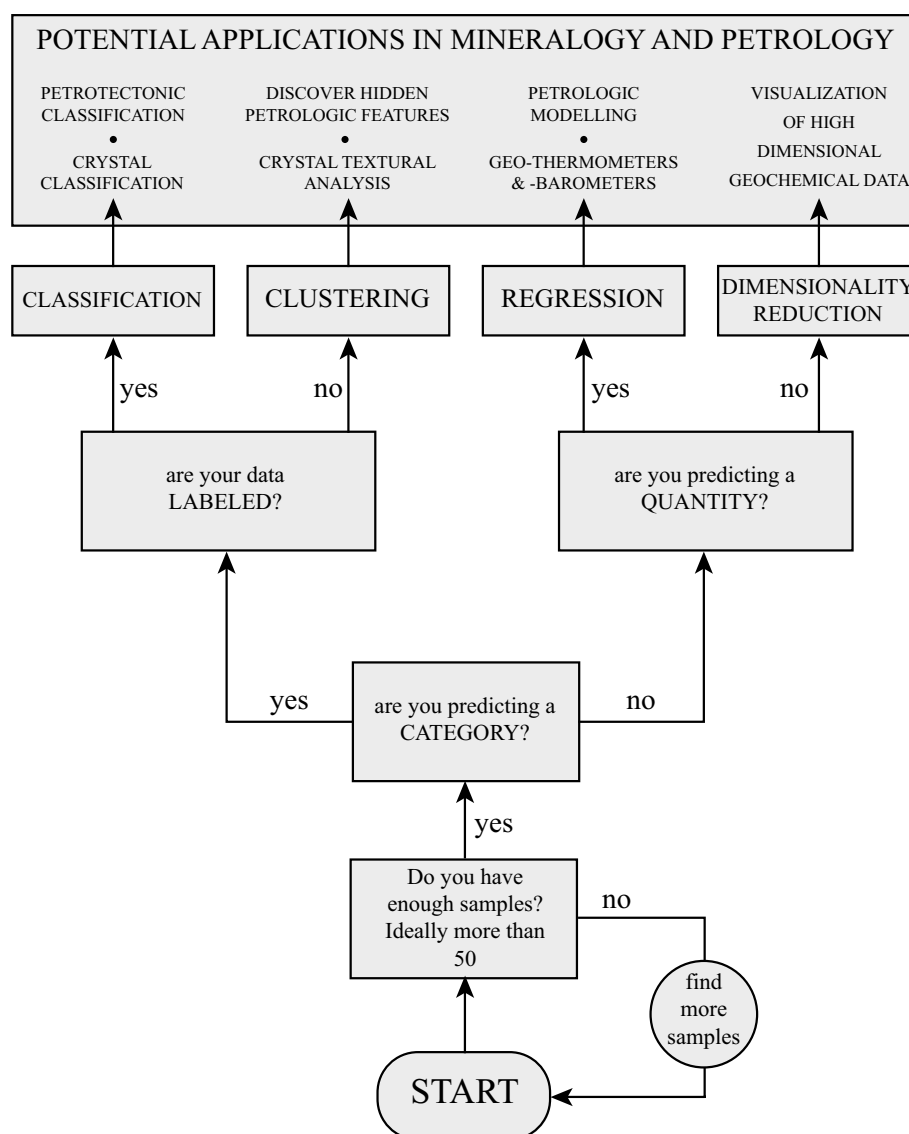
The learning process mainly consists of converting experience into "expertise" or "knowledge" (Shai and Shai 2014). Humans use past experiences to implement their learning processes. As an example, a child starts learning the alphabet by looking at any place where he finds a word or a phrase (a book, a web page, etc.). Then, typically at school, he is taught the meaning of the letters and how to combine them. As a consequence, the learning process for a child is a combination of experiences and teaching. Likewise, the experience for a learning algorithm is the training data and the output is an expertise, which usually takes the form of another computer program and/or a model that can perform a specific task (Shai and Shai 2014).

The previous example allows us to introduce the two main categories of learning in ML: (a) unsupervised and (b) supervised learning.

In unsupervised applications (such as the case of a child that starts looking at words and letters), the training dataset consists of a number of input vectors without any corresponding target values. On the contrary, in supervised applications (a child learning at school), the training dataset is labeled, meaning that the algorithm learns through examples (Bishop 2007).

Figure 1 is a flowchart illustrating the main application fields of ML (classification, clustering, regression and dimensionality reduction) and their potential use to solve representative mineralogical and petrological problems. As reported in Fig. 1, a prerequisite for a successful application of a ML technique is the availability of a suitable number of samples (indicatively more than 50). The goal is to define the right ML field (i.e., classification, clustering, regression or dimensionality reduction) in which a problem

Fig. 1 Flowchart showing the main application fields of machine learning (classification, clustering, regression and dimensionality reduction) and their applicability to representative petrological and mineralogical problems. The figure also shows how to select a specific application field in the attempt to solve a petrological or mineralogical problem

POTENTIAL APPLICATIONS IN MINERALOGY AND PETROLOGY

PETROTECTONIC CLASSIFICATION
•
CRYSTAL CLASSIFICATION

DISCOVER HIDDEN PETROLOGIC FEATURES
•
CRYSTAL TEXTURAL ANALYSIS

PETROLOGIC MODELLING
•
GEO-THERMOMETERS & -BAROMETERS

VISUALIZATION OF HIGH DIMENSIONAL GEOCHEMICAL DATA

CLASSIFICATION   CLUSTERING   REGRESSION   DIMENSIONALITY REDUCTION

yes   no   yes   no

are your data LABELED?   are you predicting a QUANTITY?

yes   are you predicting a CATEGORY?   no

yes

Do you have enough samples? Ideally more than 50   no   find more samples

START

can be treated (Fig. 1). This involves a series of decisions regarding the nature of problem.

If the problem involves categories, the first step is to select between labeled and unlabeled data. In the case of labeled data, the learning process is supervised and we are dealing with a "classification" problem (Kotsiantis 2007). Potential examples of classification problems in petrology and mineralogy are the classification of crystals in complex systems (e.g., Fischer et al. 2006) and, in the case of this study, the petro-tectonic classification using geochemical data (Pearce et al. 1984). In the case of unlabeled data, we are dealing with a "clustering" problem (Jain et al. 1999). The field of clustering is not completely unexplored in petrology (e.g., Le Maitre 1982). As an example Le Maitre (1982), discussed the basics of clustering in petrology. Potential applications of clustering problems in mineralogy and petrology are the discovery of hidden petrological

structures in geochemical data or the quantitative analysis of crystal textures (Lach-hab et al. 2010).

If the problem does not involve categories, the next step is to define whether a quantity must be predicted. If the answer is yes, we are in the field of "regression" (Smola and Schölkopf 2004). A potential petrological application of ML regression is the fitting of empirical data from experimental petrology when the mathematical formulation of the problem is not known a priori. Example applications are geothermometric and geobarometric studies or the modeling of evolutionary processes in igneous petrology. Finally, if the aim of the problem does not deal with the prediction of a quantity, we are in the field of "dimensionality reduction" (Lee and Verleysen 2009). This field is particularly useful, for example, in the context of visualization of high-dimensional petrological and mineralogical data. Example applications in petrology are the concepts of

composition and reaction spaces introduced by Thompson to solve metamorphic problems (Thompson 1982a, b).

Support vector machines (SVM)

Before presenting our analyses, we report some basic information about support vector machines. The readers interested in the details of the SVM theory and numerical methods can find full methodological descriptions in Cortes and Vapnik (1995). A summary of the mathematics behind SVMs is also given in Appendix A.

SVMs are a set of ML algorithms that are particularly useful in the context of classification (Cortes and Vapnik 1995). During the learning phase, sets of known and categorized training examples are analyzed by the SVM algorithm. The SVM then elaborates a model and assigns unknown samples to different categories (Cortes and Vapnik 1995).

The main strength points of SVMs are: (1) SVMs are effective in high-dimensional spaces; (2) SVMs can model complex and real-world problems; (3) SVMs perform well on datasets with many attributes, despite the possible low number of cases which might be available to train the model (Cortes and Vapnik 1995; Yu et al. 2005). All these features make SVMs potentially very useful in the resolution of high-dimensional petrological and geochemical problems of classification.

SVMs numerically implement the following idea: inputs are mapped to a very high-dimension feature space where a decision surface is then constructed (Cortes and Vapnik 1995). In the simplest implementation, the decision surface is linear. It consists of a hyper-plane or set of hyper-planes, which can be used for classification, regression or other tasks. The simplest way to separate two groups of data is by using a straight line (two dimensions; two chemical elements in our case, and as in the case of binary classification diagrams), a flat plane (three dimensions, three chemical elements) or an N-dimensional hyper-plane (i.e., N chemical elements). However, certain problems require a nonlinear trend to separate the groups more efficiently. SVMs handle these occurrences using nonlinear kernel functions (see "Appendix A" for further details).

SVMs have been originally developed for binary classification problems, where the algorithm learns and performs the classification of unknowns between two classes. However, most of the SVMs applications deal with problems which have a larger number of classes. These applications, such as the case of the discrimination of magmas from different tectonic environments using geochemical data, are defined as multiclass classification problems. In multiclass problems, two main strategies can be implemented: (1) One Vs One (OVO) and (2) One Vs Rest (OVR). In the OVO approach, each population is compared with each other

population, separately. The OVR approach compares each population with all the other populations simultaneously. Further details about the OVO and OVR approaches are detailed in Appendix A.

Support vector machines implementation

To evaluate the use of SVMs in the context of tectonic classification using geochemical data, we used the scikit-learn (Pedregosa et al. 2011) implementation of SVM. Scikit-learn is a Python module integrating a wide range of state-of-the-art machine-learning algorithms for medium-scale supervised and unsupervised problems. This package focuses on bringing machine learning to nonspecialists using a general-purpose high-level language (Pedregosa et al. 2011). We selected to use the scikit-learn package as this will allow other users to easily replicate our results. In addition, it represents a powerful framework for the solution of petrological and mineralogical problems in fields of clustering, regression, dimensionality reduction and, with regards to the presented case study, classification (Fig. 1). To evaluate the best strategy for the learning stage of the proposed case study, we tested both linear and nonlinear kernels using both the OVO and OVR approaches. As nonlinear kernel, we selected the so-called radial basis function (RBF) which is one of the most widely used and performing nonlinear kernels (Scholkopf et al. 1997).

Input data

For comparison, we used the same data sources used in the work of Li et al. (2015). In particular, data are retrieved from the two most comprehensive global petrological databases available to date: PetDB (http://www.earthchem.org/petdb) and GEOROC (http://georoc.mpch-mainz.gwdg.de/georoc/). In order to test SVMs as a general classification tool, we did not limit the present study to basalts but we used all volcanic rock samples for which the whole geochemical characterization of major elements (SiO_2, TiO_2, Al_2O_3, Fe_2O_{3T}, CaO, MgO, Na_2O, K_2O), selected trace elements (Sr, Ba, Rb, Zr, Nb, La, Ce, Nd, Hf, Sm, Gd, Y, Yb, Lu, Ta, Th) and isotopes ($^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{208}Pb/^{204}Pb$, $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$) were available.

Compositions characterized by SiO_2 content, on volatile free basis, ranging from 40 to 80 wt% were selected. Samples marked as altered in the databases were not considered. In addition, we consider the following tectonic environments (Fig. 2), in accordance with Frisch et al. (2011): continental arcs (CA), island arcs (IA), intra-oceanic arcs (IOA), back-arc basins (BAB), continental floods (CF), mid-ocean ridges (MOR), oceanic plateaus (OP) and ocean islands (OI) (Table 1). The entire analyzed dataset consists of a total of 3095 samples. Table 1 reports the
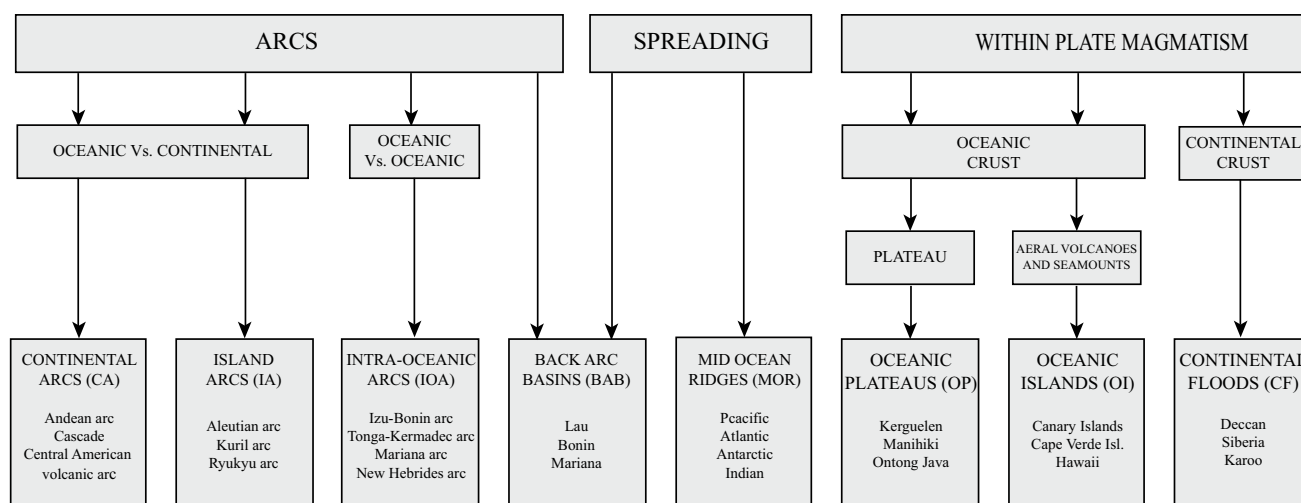
Fig. 2 Definition of the considered tectonic environments in accordance with Frisch et al. (2011). As a reference, typical magmatic regions are also indicated

number of samples from each tectonic setting as well as two statistical indicators (geometric mean and standard deviation, Table 1) for some key geochemical parameters typically used in "conventional" discrimination diagrams. To better visualize the statistical distribution of some key parameters, Fig. 3 reports the histogram distributions for SiO_2, Total Alkalies, Zr, La and Y for the different tectonic environments considered in our study. Observing both Table 1 and Fig. 3, it emerges that all the reported parameters show large variations (large standard deviations) and significant overlapping areas among the different tectonic settings, in agreement with the results reported by Li et al. (2015).

Data standardization

Standardization of the dataset is a common requirement for many machine-learning estimators. It involves in reporting each individual feature (i.e., element composition) to a standard, normally distributed population (e.g., Gaussian with zero mean and unit variance). The radial basis function used as nonlinear kernel in the present study is one of the estimators requiring a standard, normally distributed dataset (Pedregosa et al. 2011). The standardization process is shown in Fig. 4 using Sr as a representative element. Figure 4A shows the original, nonstandardized, Sr distribution. The Sr composition ranges from 0 to c.a 2800 µg/g, and the histogram is characterized by an asymmetrical shape (Fig. 4A). To standardize the Sr distribution, we first applied the Box–Cox transformation (Box and Cox 1964; Fig. 4B). Then, the obtained distribution was further transformed by removing the mean and scaling to unit variance (e.g., Templ et al. 2008; Fig. 4C).

Experiments

In order to evaluate the classification capabilities of SVMs when dealing with geochemical data, several experiments were performed using major elements, followed by trace elements and isotopes, separately. The major elements were then progressively combined with the trace element and isotopic data. Three sets of experiments have been performed. For the first two experimental sets, the entire dataset was split into two groups containing 70 % (learning population) and 30 % (test population) of samples. Samples were randomly assigned to the learning and test populations. The learning population was then used to train the ML algorithms, and the test population was analyzed as unknown. We first evaluated the learning capabilities of the OVO and OVR strategies using a linear kernel. Successively, we evaluated the performances of a nonlinear kernel. Finally, we evaluated the resulting best strategy on the whole dataset using a leave-one-out approach (LOO; James et al. 2013). LOO is one of the simplest cross validation methods and consists in learning the system by taking all the samples except one. The sample which is left out is then introduced within the system as unknown. Thus, for n samples, we performed n different trainings, one for each unknown sample.

Metrics

We define the classification score (CS) as the ability of a specific SVM algorithm to discriminate among the different tectonic environments as defined by Frisch et al. (2011) and queried in the reference databases (i.e., GEOROC and PetDB; Fig. 2). CS is quantitatively defined as the ratio

Table 1 Synoptic table for the considered tectonic environments

Tectonic setting	No. of samples	SiO$_2$ (wt%)	1 sd	Total Alkalies (wt%)	1 sd	Mg#	1 sd	Zr (µg/g)	1 sd	Nb (µg/g)	1 sd	La (µg/g)	1 sd	Y (µg/g)	1 sd	Compositions
Continental arcs (CA)	840	55	6	5	1	51	9	146	76	9	12	21	20	19	8	Basaltic to dacitic
Intra-oceanic arcs (IOA)	654	54	5	3	1	47	12	56	45	1	4	4	8	23	13	Mostly basaltic and andesitic, subordinate dacites
Island arcs (IA)	266	55	7	5	1	46	10	105	63	3	11	12	8	26	9	Mostly basaltic and andesitic
Back-arc basins (BAB)	123	53	4	3	1	47	13	82	80	2	4	4	12	26	11	Basaltic, subordinate andesites and dacites
Continental floods (CF)	258	51	6	3	2	51	15	162	152	13	19	19	21	28	15	Basaltic, subordinate andesites
Mid-ocean ridges (MOR)	394	50	3	3	1	55	10	110	147	6	22	6	13	33	20	Mostly basaltic
Oceanic plateaus (OP)	58	54	8	4	3	43	21	143	440	13	44	13	37	31	34	Basaltic, subordinate andesites and dacites
Ocean Islands (OI)	502	48	4	3	3	57	13	188	175	23	38	20	27	27	10	Basaltic, subordinate andesites

The number of samples utilized in our study for each environment, the geometric mean and the standard deviation (1 sd) of some key parameters (SiO$_2$, Total Alkalies, Mg#, Zr, Nb, La, Y) are shown

between the number of correctly discriminated samples and the total number of samples of the test population. A CS equal to 1 means that the algorithm is capable to recognize and classify all the samples belonging to the test population to the relative tectonic environment (Fig. 2). A CS equal to 0 means that the algorithm is unable to allocate any sample of the test population to the relative tectonic environment.

Results and discussion

As reported above, in the first sets of experiments, a comparison was made between classification capabilities of the OVO and OVR strategies using a linear kernel for major elements (Fig. 5A), trace elements (Fig. 5B), isotopes (Fig. 5C) and the combination of major plus trace elements and isotopes (Fig. 5D) as input variables. In all cases reported in Fig. 5, the CS increases as the number of dimensions of the system increases (i.e., as the number of considered geochemical elements increases). Considering only major elements (Fig. 5A), the classification scores are in range from 0.45 (2D) to 0.66 (8D) and 0.47 (2D) to 0.69 (8D) for the OVR and OVO approaches, respectively. Figure 5B reports the CS obtained considering only trace elements as input parameters, plotted against the dimension of the system. We reiterate that the dimension of the problem is the number of input parameters and therefore, the number of geochemical elements evaluated by the system. Classification scores for trace elements (Fig. 5B) range from 0.43 (2D) to 0.73 (16D) and from 0.43 (2D) to 0.78 (16D) for the OVR and OVO approach, respectively. Classification scores considering only isotopes (Fig. 5C) range from 0.30 (2D) to 0.45 (5D) and from 0.32 (2D) to 0.47 (5D) for the OVR and OVO approach, respectively. Combining major elements, trace elements and isotopes (Fig. 5D), the classification score increases progressively reaching values up to 0.84 (29D) and 0.88 (29D) for the OVR and OVO methods, respectively. It is notable that the OVO approach consistently shows higher classification scores relative to the OVR method, evidencing that the OVO approach is more efficient and therefore more suitable than OVR in the classification of the considered samples. We note that the results are in agreement with those reported by Hsu and Lin (2002), who suggest that OVO methods are more suitable for practical uses relative to other methods. This is due to the fact that, although the OVR approach is faster than OVO, it experiences a bias due to the lack of information regarding the boundaries defining each single population. This does not happen with the OVO approach which, in most cases, provides better results.

At the second stage of our analysis, we evaluated whether the use of a nonlinear kernel was able to improve the learning capabilities of the system compared to the linear one. This was achieved by comparing the results
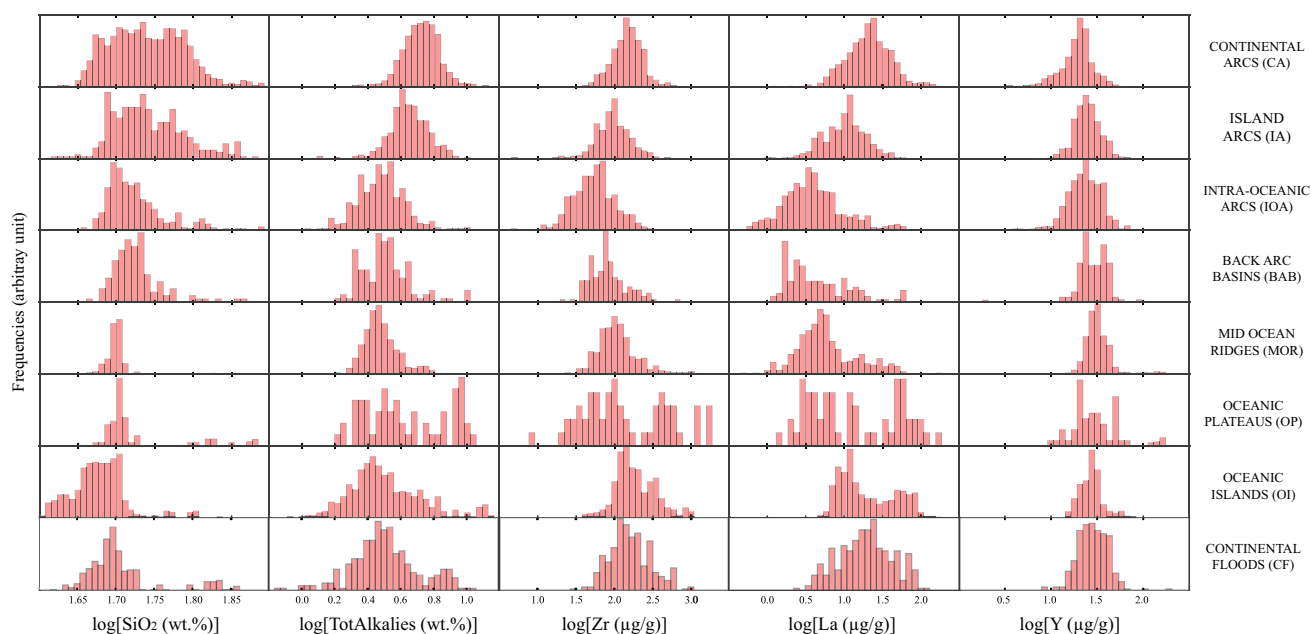
Fig. 3 Histograms for selected key parameters (SiO$_2$, Total Alkalies, Zr, La and Y) for the different tectonic environments: continental arcs (CA), island arcs (IA), intra-oceanic arcs (IOA), back-arc basins (BAB), continental floods (CF), mid-ocean ridges (MOR), oceanic plateaus (OP) and ocean islands (OI)

obtained using the RBF kernel function with those obtained from the previous set of experiments, where a linear kernel was utilized. Based on the results of the first experimental set reported above, the OVO approach was used for this second set of experiments. Results are reported in Fig. 6. In this case, classification scores range from 0.53 (2D) to 0.79 (8D), 0.46 (2D) to 0.87 (16D), 0.40 (2D) to 0.79 (5D) and 0.53 (2D) to 0.93 (29D) for major elements (Fig. 6A), trace elements (Fig. 6B), isotopes (Fig. 6C) and the combination of major + trace elements + isotopes (Fig. 6D), respectively.

These results are consistently superior to those obtained for the linear kernel (Fig. 5A–C). To aid in the understanding of the result, those obtained using the linear kernel are reported as a reference in Fig. 6. It is clear that the highest classification score (CS = 0.93) is obtained by coupling all the major and trace elements with isotopic data (29D).

According to the above results, we elected to perform a third set of experiments investigating the classification capabilities of the most performing configuration (29D + nonlinear kernel + OVO method) in the attempt to determine the tectonic setting of each sample within the dataset. For the third set of experiments, the LOO approach was employed. As reported above, it involves learning the system by taking all the samples except one. The sample which is left out is then introduced to the system as an unknown. The procedure is repeated for all the samples belonging to the input dataset, and the performance of the experiment is evaluated.

Results are reported in Fig. 7 in the form of a "confusion matrix". This graphical representation is a useful method to display information about actual and predicted classifications allowing a straightforward visualization of results (Provost and Kohavi 1998). In a confusion matrix, each column represents the instances in a predicted class, whereas each row represents the instances in an actual class. As a consequence, correct estimations are reported in the cells belonging to the main diagonal of the matrix; the errors are reported in the other cells. Classification scores higher than 84 % are obtained for all the studied tectonic settings with the only exception of BAB samples, for which a score of 65 % is obtained (Fig. 7). Classification scores for samples belonging to six tectonic settings are better than 89.5 %. They are CA, IOA, IA, CF, OI and MOR characterized by classification scores of 97.3, 92.0, 89.5, 95.0, 99.2 and 92.4 %, respectively. OP shows a lower classification score, but remains above 84 %. It is interesting to note that these results are superior to those reported by (Snow 2006) for IA, OI and MOR environments of about 83, 75 and 76 %, respectively.

The case of BAB is intriguing. It is characterized by the lowest classification score (65 %) among the considered tectonic settings. This means that about 35 % of the samples belonging to the BAB tectonic setting have been misclassified. Among them, significant portions were classified as IOA (27.6 %) and MOR (4.1 %). Only a few samples were attributed to CA (1.6 %), IA (0.8 %), CF (0.8 %), OP (<0.5 %) and OI (<0.5 %).
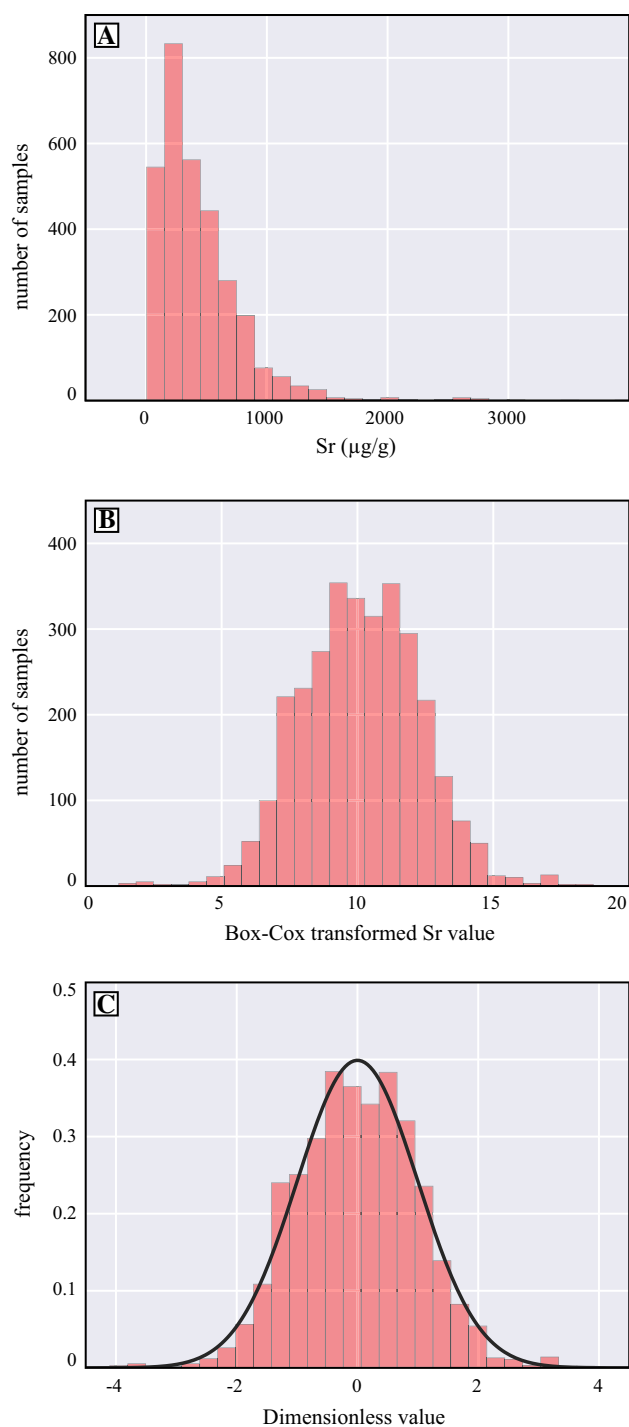
Fig. 4 Illustration of the normalization procedure using Sr as a representative element: **a** original univariate histogram distributions; **b** histogram distributions after the application of the Box–Cox transformation; **c** histogram distributions after the subtraction of the mean and the transformation to get a variance equal to 1

Although the percentage of success in classifying BAB samples is not as large as for the other tectonic environments, noteworthy is the fact that, in the case of a

statistically representative population, BAB represents the modal value derived from the classification. We emphasize, however, that the ML system must be used for the evaluation of a statistically representative set of unknown samples and the use of single samples should be avoided. In fact, if one introduces a single BAB sample in the ML system, there is a 65 % of probability that it is correctly classified and 35 % that it is erroneously classified. Conversely, for example, if one uses a population of 100 BAB samples, a reasonable result might be: 65 BAB, 28 IOA, 4 MOR, 2 CA and one between IA or CF. This points to BAB as the modal value and hence to the most probable petro-tectonic association.

Noteworthy is the fact that, despite the lower percentage of success of BAB classification relative to the other tectonic environments, this represents a step forward compared to previous studies (Saccani 2015; Li et al. 2015) reporting the inability of current methods to discriminate BAB samples. The fact that the main sources of uncertainties are the spreading centers and arch environment is, however, not surprising. This is directly related to the petrological processes governing the genesis and evolution of BAB rocks. In particular, several authors (Taylor and Martinez 2003; Pearce and Stern 2006) highlighted the transitional nature of a large portion of BAB magmatic compositions lying between MOR and arc setting compositions.

The above discussion highlights the great potentials of ML methods in classifying and discriminating the tectonic setting of igneous rocks on the basis of their geochemical and isotopic data. However, although promising, these methods suffer from some limitations that need to be highlighted. In particular, the proposed ML system, as with any other data elaboration technique, must not be considered as a "magic box" where input petrological and geochemical data are transformed into a classification graph or datasheet to be immediately utilized for scientific interpretations. The method needs to be integrated with other techniques such as fieldwork, petrographic observations, classic geochemical studies and geophysical data to obtain a clear picture of the geologic framework from which samples are collected. Neglecting to do so might compromise a correct interpretation of the ML output exposing the user to some risks. One of the most important is to lose focus about the petrological processes that acted to generate the compositional variability of the samples used in the analyses, irrespective of their correct classification or not. As an example, it might be possible to discover "anomalous" trends of samples using conventional petrological diagrams, such as binary discrimination diagrams. This would allow these trends to be modeled and to advance hypotheses to explain their petrological nature. This task cannot be undertaken using the ML classification scheme proposed in our work. This does not mean, however, that ML techniques cannot be used to
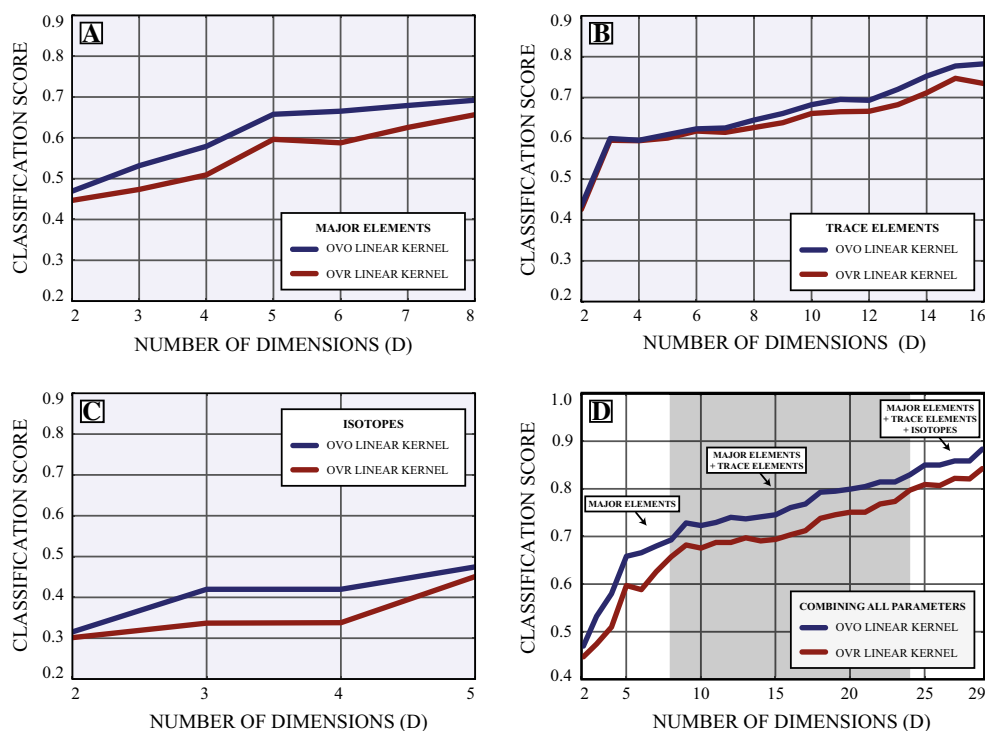
Fig. 5 Classification scores for the linear kernel plotted against the number of dimensions (*D*; i.e., number of chemical elements or isotopes) for the OVO and OVR approaches; **a** results for major elements; **b** results for trace elements; **c** results for isotopes; **d** results for the combination of major elements + trace elements + isotopes
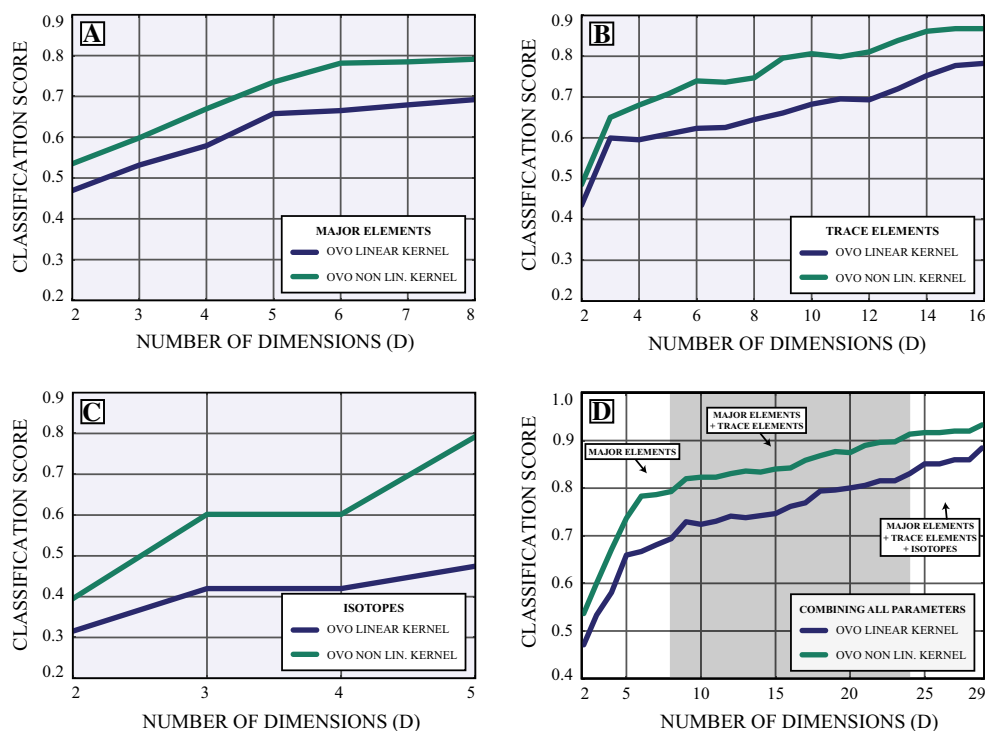
Fig. 6 Classification scores for the nonlinear kernel plotted against the number of dimensions (*D*; i.e., number of chemical elements or isotopes) using the OVO approach; **a** results for major elements; **b** results for trace elements; **c** results for isotopes; **d** results for the combination of major elements + trace elements + isotopes. Results for the linear kernel are also reported for comparison
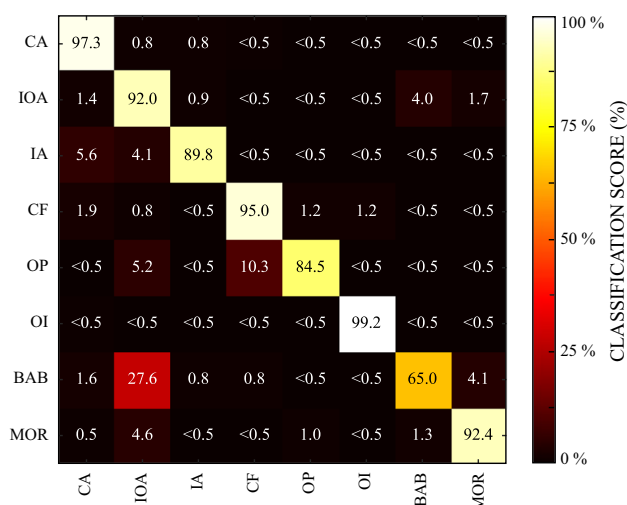
Fig. 7 Confusion matrix of the LOO cross calibration using major elements + trace elements + isotopes (29D), the nonlinear kernel and the OVO approach. See text for details

model petrological processes. As shown in Fig. 1, the qualitative and quantitative modeling of petrological processes can be achieved by ML techniques in the fields of "dimensionality reduction" and "regression", respectively.

A further limitation of the proposed classification system that must be highlighted is the inability in defining new classification groups that are different from those which are already defined. As an example, samples affected by secondary processes that altered their original geochemical composition will be necessarily classified in one of the available tectonic settings, possibly producing biases and misclassifications. In this case, preliminary petrographic inspections and geochemical investigations can be decisive in excluding those samples from the ML process. It is therefore important to combine classic petrological and geochemical approaches, together with the ML techniques, in order to resolve the multidimensional and complex nature of petrological problems.

Concluding remarks

In this contribution, we introduced the potentials of machine learning in petrology and mineralogy. We explored the applicability of support vector machines in discriminating among different tectonic environments using a large number of dimensions (up to 29D; i.e., 29 parameters, including major, trace elements and isotopes) and the two largest and comprehensive global petrological databases: PetDB and GEOROC.

We demonstrated that: (1) Machine learning has outstanding potentials in petrological and mineralogical studies; (2) support vector machines are robust and useful tools in addressing the complexity underlying the extreme compositional variability often encountered in petrological studies; (3) support vector machines are able to discriminate among different tectonic settings; (4) trace elements alone display a good discriminatory power for the different tectonic settings (CS = 0.87); (5) the combination of major elements, trace elements and isotopic data provides much more reliable results (CS = 0.93) compared to methods in which these three groups of geochemical parameters are analyzed separately (CS = 0.79 for major elements, CS = 0.87 for trace elements and CS = 0.79 for isotopes).

With regards to the case study, our model is able to discriminate the tectonic setting for studied rock samples with high percentage of success. The only exception is the dataset from back-arc basins that shows lower percentages of success. This is interpreted as a result of the transitional nature of a large proportion of back-arc basin magmatic compositions lying between mid-ocean ridge and arc setting compositions. The results of this study can have important implications in petrology opening new perspectives for petrologists and geochemists. A potential application is the study of the large amount of data arising from experimental petrology in order to find new relations and uncovered patterns. Another example is the development of more robust geothermometers and/or geobarometers by joining results deriving from different techniques. A further example of application might be the determination of the provenance of crypto-tephra using geochemical data, in order to correlate distal tephra to past volcanic events. Until now, as in the case of tectonic environment determination, geochemical tephro-chronological correlations are mainly performed using binary or ternary diagrams (Tomlinson et al. 2015). The combined use of large databases coupled with machine-learning techniques might provide a more robust quantitative approach to better correlate crypto-tephra layers to volcanic sources.

Appendix A: mathematical principles of support vector machines

An extensive introduction to the mathematical principles of support vector machines is reported in Abedi et al. (2012) and Cortes and Vapnik (1995). To introduce the

formulation of support vector machines, we first discuss a two-class problem.

Consider a training dataset of S dimensional samples (e.g., S chemical elements as input) x_i with $i = 1, 2, 3,\ldots, n$ where n is the number of samples. To each sample, a label y_i is assigned. The label y_i is equal to 1 for the first class and -1 for the second class.

In the case the two classes are linearly separable, then there exists a group of linear separators that satisfy the following equation (Kavzoglu and Colkesen 2009):

$$w \cdot x_i + b \geq +1 \quad \text{for} \quad y_i = +1$$

$$w \cdot x_i + b \leq -1 \quad \text{for} \quad y_i = -1$$

As a consequence, the separating hyper-plane can be formalized as a decision function:

$$f(x) = \text{sgn}(wx + b)$$

with sgn(x) defined as follows:

$$\text{sgn}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -1 & \text{if} \quad x < 0 \end{cases}$$

The parameters of w and b can be obtained by solving the optimization function:

$$\text{minimize } \tau(w) = \frac{1}{2}w^2$$

subject to:

$$y_i((wx_i) + b) \geq 1, \quad i = 1, 2, 3, \ldots, n$$

An example of two-dimensional problem where two different populations can be divided by a linear function is reported in Fig. 8A. However, there are problems where a nonlinear trend can separate the different populations more efficiently (Fig. 8B).

In these cases, a projection function $\phi(x)$ can be utilized to map the training data form the original space x to a Hilbert space X. This means that a nonlinear function is learned by a linear learning machine in a high-dimensional feature space while the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space (Cristianini and Shawe-Taylor 2000). This is called "kernel trick" and means that the kernel function transforms the data into a higher dimensional feature space allowing for performing a linear separation (Cortes and Vapnik 1995).

As reported by Abedi et al. (2012), the training algorithm in the Hilbert space only depends on data in this space through a dot product (i.e., a function with the form $\phi(x_i) \times \phi(x_j)$). As a consequence, a kernel function K can be formalized as follows:
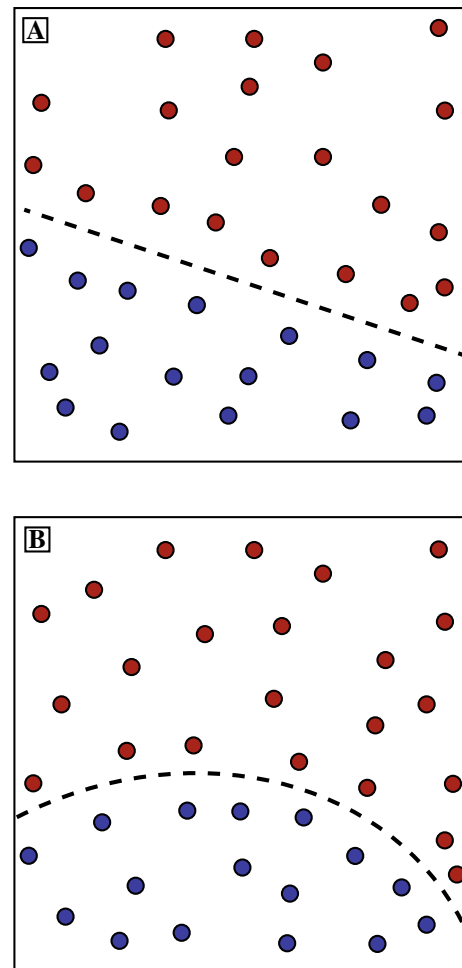
Fig. 8 Simplified examples of 2D populations that can be separated by a linear (**a**) and nonlinear (**b**) functions

$$K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$$

The two-class problem can be also solved as follows (El-Khoribi 2008):

$$\text{maximize} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{ij=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to:

$$\alpha_i \geq 0, \quad i = 1, 2, 3, \ldots, n \quad \text{and} \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

The decision function can be now rewritten as (Yang et al. 2008):

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_j) \right)$$

Fig. 9 Exemplification of the OVR and OVO approaches in 2D for the discrimination among the three populations reported in **a**. In OVO (**b**, **c**) each population is compared with each of the population separately ($M − 1$) times; in OVR (**d**) each population is compared against all the other populations mixed together
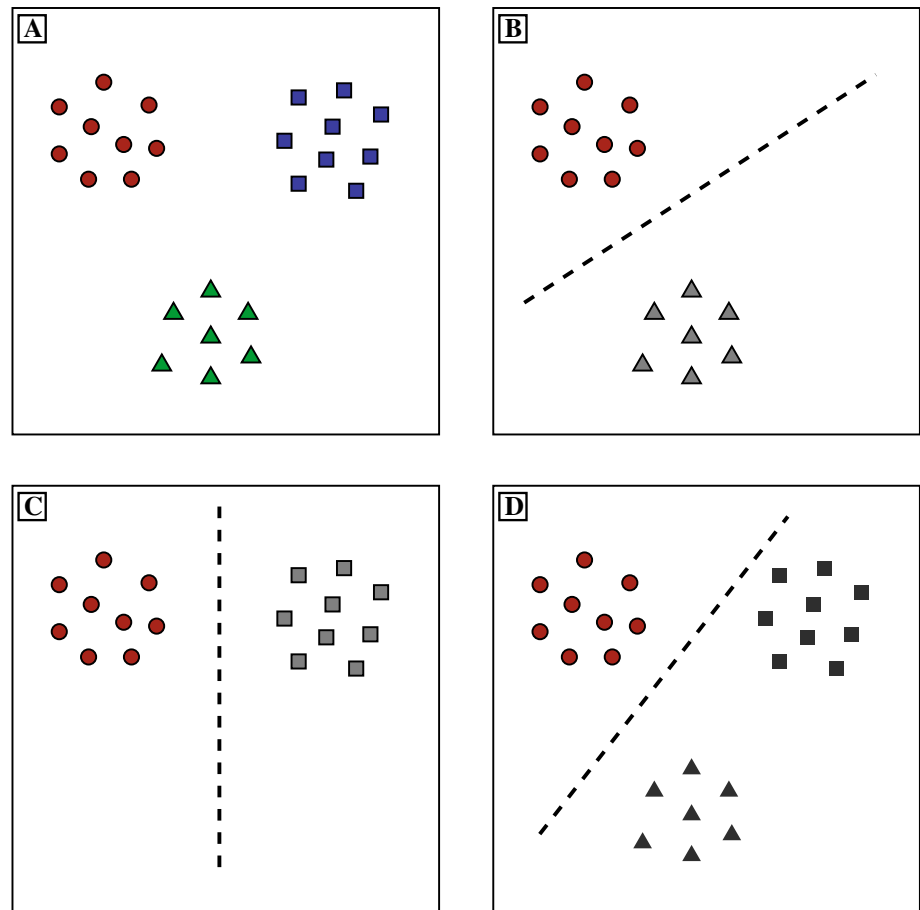
Many potential functions can be utilized as $K\left(x_i, x_j\right)$(Zuo and Carranza 2011). Among these, the radial basis function (RBF) utilized in our work is defined as follows:

$$K\left(x_i, x_j\right) = e^{-\gamma\left(x_i - x_j\right)^2}$$

As reported by Cortes and Vapnik (1995), support vector machines were originally developed for the solution of two-class problems, but many of the potential applications are characterized by more than two classes (multiclass problems). In order to solve multiclass problems, the two most popular approaches are the One Vs One (OVO) and the One Vs Rest (OVR) approach (Fig. 9). In OVO, one SVM classifier is built for all possible pairs of classes (Fig. 9B, C) (Knerr et al. 1990; Dorffner et al. 2001). The output from each classifier is obtained in the form of a class label. The class label with the highest frequency is assigned to that point in the data vector (Hsu and Lin 2002). Since the number of SVMs required in this approach is $M(M − 1)/2$, it is not suitable for those datasets characterized by a large number of classes (Dorffner et al. 2001).

On the contrary, in OVR, one SVM is built for each of the M classes. The SVM for a particular class is constructed using the training examples from that class as positive examples and the training examples of the rest of (M-1) class as negative examples (Fig. 9D).

In other words, in the OVO (One Vs One) approach, each population is compared with each other population, separately. In the OVR (One Vs Rest) approach, each population is compared with all the other populations mixed together, simultaneously.

Appendix B: the logic behind classification

Figure 10 reports a flowchart showing the steps to be implemented to determine the tectonic environment of igneous. The first step consists in verifying whether the learning process has been already performed. In the case the learning is missing, a new learning process is required. To complete this task, the reference dataset has to be normalized and split into two portions: the learning and test dataset. The role of the learning dataset is to train the system and develop a provisional model. The role of the test dataset is to check the goodness of the provisional model developed using the learning dataset. To complete this task, the samples
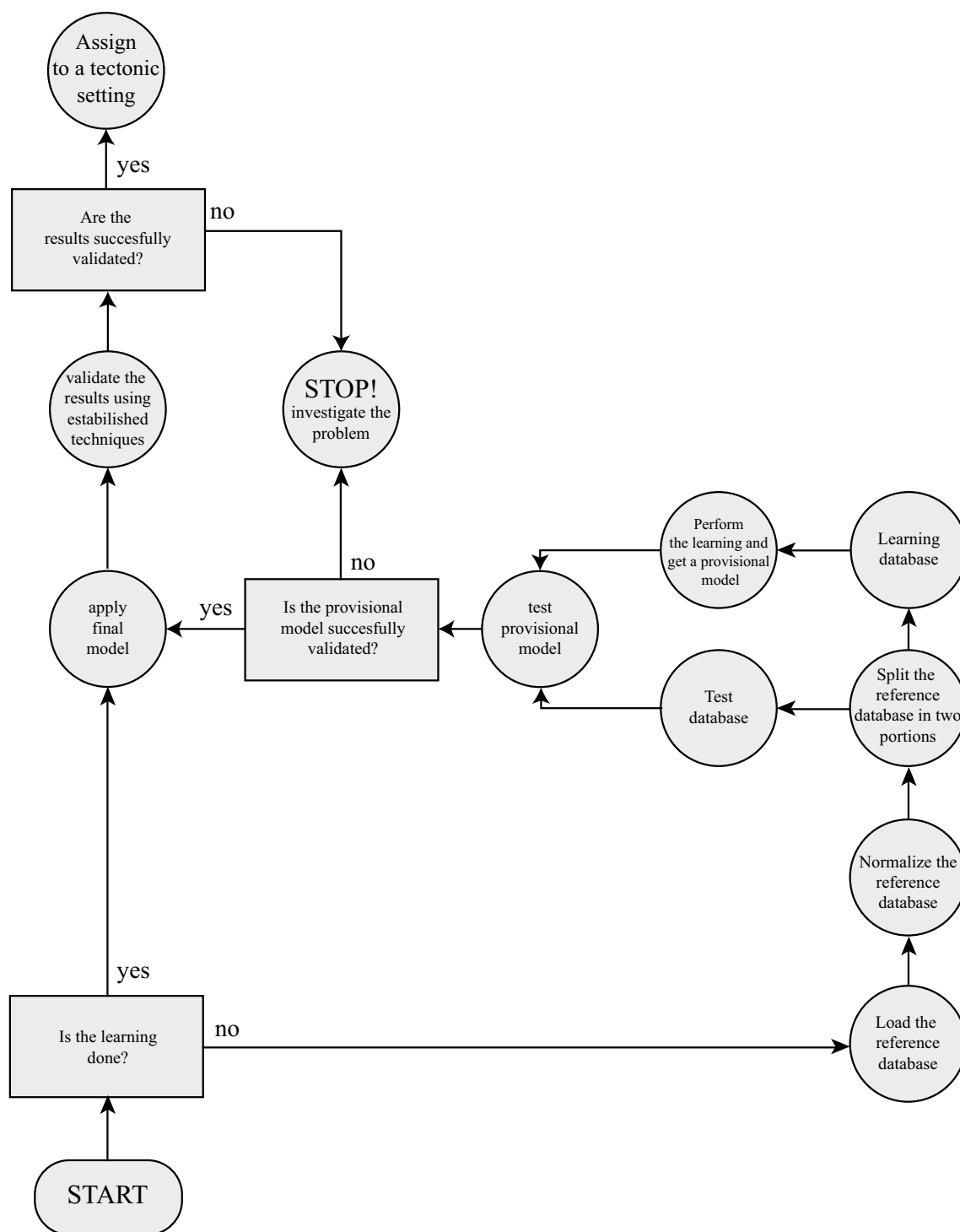
Fig. 10 Flowchart showing the steps to be performed in the attempt of determining the tectonic setting of unknown samples

belonging to the test dataset are evaluated as unknowns using the provisional model. If the validation process is completed successfully, the provisional model is converted to a final model. On the contrary, the whole classification process is aborted and more detailed studies are required.

When the final model is ready, the samples belonging to the unknown population are processed by the system. Results are then cross-validated using conventional techniques such as petrographic inspections, classical geochemical investigations and field observations.

Finally, if the provisional results are confirmed, the unknown samples can be safely assigned to a specific tectonic setting. On the contrary, further investigations are needed.

References

Abedi M, Norouzi G-H, Bahroudi A (2012) Support vector machine for multi-classification of mineral prospectivity areas. Comput Geosci 46:272–283. doi:10.1016/j.cageo.2011.12.014

Agrawal S, Guevara M, Verma SP (2004) Discriminant analysis applied to establish major-element field boundaries for tectonic varieties of basic rocks. Int Geol Rev 46:575–594. doi:10.2747/0020-6814.46.7.575

Agrawal S, Guevara M, Verma SP (2008) Tectonic discrimination of basic and ultrabasic volcanic rocks through log-transformed ratios of immobile trace elements. Int Geol Rev 50:1057–1079. doi:10.2747/0020-6814.50.12.1057

Bishop C (2007) Pattern recognition and machine learning. Springer, New York

Box GEP, Cox DR (1964) An analysis of transformations. J Roy Stat Soc B Met 26:211–252. doi:10.2307/2287791

Cannata A, Montalto P, Aliotta M et al (2011) Clustering and classification of infrasonic events at Mount Etna using pattern recognition techniques. Geophys J Int 185:253–264. doi:10.1111/j.1365-246X.2011.04951.x

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. doi:10.1007/BF00994018

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge

Dorffner G, Bischof H, Hornik K (2001) Artificial neural networks—ICANN 2001. Springer, Berlin

El-Khoribi RA (2008) Support vector machine training of HMT models for land cover image classification. ICGST-GVIP 8:7–11

Fischer CC, Tibbetts KJ, Morgan D, Ceder G (2006) Predicting crystal structure by merging data mining with quantum mechanics. Nat Mater 5:641–646. doi:10.1038/nmat1691

Frisch W, Meschede M, Blakey R (2011) Plate tectonics. Continental drift and mountain building. Springer, New York

Goldstein EB, Coco G (2014) A machine learning approach for the prediction of settling velocity. Water Resour Res 50:3595–3601. doi:10.1002/2013WR015116

Grimes CB, Wooden JL, Cheadle MJ, John BE (2015) "Fingerprinting" tectono-magmatic provenance using trace elements in igneous zircon. Contrib Miner Petrol 170:46. doi:10.1007/s00410-015-1199-3

Hsu C-WC, Lin CC-J (2002) A comparison of methods for multiclass support vector machines. Neural Netw IEEE Trans 13:415–425. doi:10.1109/72.991427

Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. Int J Remote Sens 23:725–749. doi:10.1080/01431160110040323

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv (CSUR) 31:264–323. doi:10.1145/331499.331504

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349:255–260. doi:10.1126/science.aaa8415

Kavzoglu T, Colkesen I (2009) A kernel functions analysis for support vector machines for land cover classification. Int J Appl Earth Obs Geoinf 11:352–359. doi:10.1016/j.jag.2009.06.002

Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Neurocomputing: algorithms, architectures and applications. Springer, Berlin, pp 41–50

Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. Informatica 31:249–268

Lach-hab M, Yang S, Vaisman II, Blaisten-Barojas E (2010) Novel approach for clustering zeolite crystal structures. Mol Inform 29:297–301. doi:10.1002/minf.200900072

Le Maitre RW (1982) Numerical petrology: statistical interpretation of geochemical data. Elsevier, Amsterdam

Lee JA, Verleysen M (2009) Quality assessment of dimensionality reduction: Rank-based criteria. Neurocomputing 72:1431–1443. doi:10.1016/j.neucom.2008.12.017

Li C, Arndt NT, Tang Q, Ripley EM (2015) Trace element indiscrimination diagrams. Lithos 232:76–83. doi:10.1016/j.lithos.2015.06.022

Masotti M, Falsaperla S, Langer H et al (2006) Application of support vector machine to the classification of volcanic tremor at Etna, Italy. Geophys Res Lett 33:L20304. doi:10.1029/2006GL027441

Meschede M (1986) A method of discriminating between different types of mid-ocean ridge basalts and continental tholeiites with the Nb–Zr–Y diagram. Chem Geol 56:207–218. doi:10.1016/0009-2541(86)90004-5

Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press, Cambridge

Pearce JA (1976) Statistical analysis of major element patterns in basalts. J Petrol 17:15–43. doi:10.1093/petrology/17.1.15

Pearce JA, Cann JR (1973) Tectonic setting of basic volcanic rocks determined using trace element analyses. Earth Planet Sci Lett 19:290–300. doi:10.1016/0012-821X(73)90129-5

Pearce JA, Norry MJ (1979) Petrogenetic implications of Ti, Zr, Y, and Nb variations in volcanic rocks. Contrib Miner Petrol 69:33–47. doi:10.1007/BF00375192

Pearce J, Stern R (2006) Origin of back-arc basin magmas: trace element and isotope perspectives. In: Christie DM, Fisher CR, Lee SM, Givens S (eds) Back-arc spreading systems: geological, biological, chemical, and physical interactions. American Geophysical Union, Washington, DC

Pearce JA, Harris NBW, Tindle AG (1984) Trace element discrimination diagrams for the tectonic interpretation of granitic rocks. J Petrol 25:956–983. doi:10.1093/petrology/25.4.956

Pedregosa F, Varoquaux GG, Gramfort A et al (2011) Scikit-learn: machine Learning in Python. J Mach Learn Res 12:2825–2830

Petrelli M, Perugini D, Moroni B, Poli G (2003) Determination of travertine provenance from ancient buildings using self-organizing maps and fuzzy logic. Appl Artif Intell 17:885–900. doi:10.1080/713827251

Provost F, Kohavi R (1998) Guest editors' introduction: on applied research in machine learning. Mach Learn 30:127–132. doi:10.1023/A:1007442505281

Saccani E (2015) A new method of discriminating different types of post-Archean ophiolitic basalts and their tectonic significance using Th–Nb and Ce–Dy–Yb systematics. Geosci Front 6:481–501. doi:10.1016/j.gsf.2014.03.006

Scholkopf B, Burges CJC, Girosi F et al (1997) Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans Signal Process 45:2758–2765. doi:10.1109/78.650102

Shai S-S, Shai B-D (2014) Understanding machine learning: from theory to algorithms. Cambridge University Press, Cambridge

Shervais JW (1982) Ti-V plots and the petrogenesis of modern and ophiolitic lavas. Earth Planet Sci Lett 59:101–118. doi:10.1016/0012-821X(82)90120-0

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222. doi:10.1023/B:STCO.0000035301.49549.88

Snow CA (2006) A reevaluation of tectonic discrimination diagrams and a new probabilistic approach using large geochemical databases: moving beyond binary and ternary plots. J Geophys Res 111:B06206. doi:10.1029/2005JB003799

Taylor B, Martinez F (2003) Back-arc basin basalt systematics. Earth Planet Sci Lett 210:481–497. doi:10.1016/S0012-821X(03)00167-5

Templ M, Filzmoser P, Reimann C (2008) Cluster analysis applied to regional geochemical data: problems and possibilities. Appl Geochem 23:2198–2213. doi:10.1016/j.apgeochem.2008.03.004

Thompson JB (1982a) Composition space; an algebraic and geometric approach. Rev Miner Geochem 10:1–31

Thompson JB (1982b) Reaction space: an algebraic and geometric approach. Rev Miner Geochem 10:33–52

Tomlinson E, Smith V, Albert P (2015) The major and trace element glass compositions of the productive Mediterranean volcanic sources: tools for correlating distal tephra layers in and around Europe. Quat Sci Rev 118:48–66. doi:10.1016/j.quascirev.2014.10.028

Verma SP, Pandarinath K, Verma SK, Agrawal S (2013) Fifteen new discriminant-function-based multi-dimensional robust diagrams for acid rocks and their application to Precambrian rocks. Lithos 168–169:113–123. doi:10.1016/j.lithos.2013.01.014

Vermeesch P (2006a) Tectonic discrimination diagrams revisited. Geochem Geophys Geosyst. doi:10.1029/2005GC001092

Vermeesch P (2006b) Tectonic discrimination of basalts with classification trees. Geochim Cosmochim Acta 70:1839–1848. doi:10.1016/j.gca.2005.12.016

Wood DA (1980) The application of a Th-Hf-Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary Volcanic Province. Earth Planet Sci Lett 50:11–30. doi:10.1016/0012-821X(80)90116-8

Yang Q, Li X, Shi X (2008) Cellular automata for simulating land use changes based on support vector machines. Comput Geosci 34:592–602. doi:10.1016/j.cageo.2007.08.003

Yu H, Yang J, Han J, Li X (2005) Making SVMs scalable to large data sets using hierarchical cluster indexing. Data Min Knowl Discov 11:295–321. doi:10.1007/s10618-005-0005-7

Zuo R, Carranza EJM (2011) Support vector machine: a tool for mapping mineral prospectivity. Comput Geosci 37:1967–1975. doi:10.1016/j.cageo.2010.09.014

Terms and Conditions